

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
8 July 2004 (08.07.2004)

PCT

(10) International Publication Number
WO 2004/056982 A2

(51) International Patent Classification⁷:

C12N 9/00

(21) International Application Number:

PCT/EP2003/014674

(22) International Filing Date:

19 December 2003 (19.12.2003)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

0230014.3 23 December 2002 (23.12.2002) GB
60/436,380 23 December 2002 (23.12.2002) US

(71) Applicant (for all designated States except US): DEVGEN NV [BE/BE]; Technologiepark 30, B-9052 Gent-Zwijnaarde (BE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): DE WILDE, Gert, Jules, Hector [BE/BE]; Dr. Armand Rubbensstraat 25, B-9240 Zele (BE).

(74) Agents: MURPHY, Colm, Damien et al.; Boult Wade Tennant, 70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 2004/056982 A2

(54) Title: KINASE SEQUENCES USEFUL FOR DEVELOPING COMPOUNDS FOR THE PREVENTION AND/OR TREATMENT OF METABOLIC DISEASES AND NUCLEOTIDE SEQUENCES ENCODING SUCH KINASE SEQUENCES

(57) Abstract: The present invention relates to nucleotide sequences that encode and may be used to express amino acid sequences that are useful in the identification and development of compounds with activity as pharmaceuticals, in particular of compounds for the prevention and treatment of metabolic diseases such as diabetes and obesity. The invention also relates to the amino acid sequences - such as proteins and polypeptides - that are encoded by, and that may be obtained by suitable expression of, the nucleotide sequences of the invention, particularly the amino acid sequences of J1K, PSK, TAO1 and Q9P2I6. The invention also relates to various uses and modulators of, and methods incorporating, the nucleotide and amino acid sequences of the invention.

Kinase sequences useful for developing compounds for the prevention and/or treatment of metabolic diseases and nucleotide sequences encoding such kinase sequences.

The present invention relates to nucleotide sequences that are useful in the 5 pharmaceutical field.

In particular, the invention relates to nucleotide sequences that encode and/or may be used to express amino acid sequences that are useful in the identification and/or development of compounds with (potential) activity as pharmaceuticals, in particular of compounds for the prevention and/or treatment of metabolic diseases such as diabetes 10 and obesity. These nucleotide sequences, which will be further described below, will also be referred to herein as "*nucleotide sequences of the invention*".

The invention also relates to the amino acid sequences - such as proteins and/or polypeptides - that are encoded by, and/or that may be obtained by suitable expression of, 15 the nucleotide sequences of the invention. These amino acid sequences, which will be further described below, will also be referred to herein as "*amino acid sequences of the invention*".

The invention also relates to the use of the nucleotide sequences of the invention, preferably in the form of a suitable genetic construct as described below, in the transformation of host cells and/or host organisms, for example for the expression of the 20 amino acid sequences of the invention. The invention also relates to host cells and/or host organisms that have been transformed with the nucleotide sequences of the invention and/or that can express the amino acid sequences of the invention.

The invention further relates to methods for the identification and/or development of compounds that can modulate the (biological) activity of the amino acid sequences of 25 the invention, in which the abovementioned nucleotide sequences, amino acid sequences, genetic constructs, host cells and/or host organisms are used. Such methods which will usually be in the form of an assay or screen, as will also be further described below.

The invention also relates to the use of the nucleotide sequences, amino acid sequences, genetic constructs, host cells and/or host organisms of the invention in 30 (methods for) identifying and/or developing compounds that can modulate the (biological) activity of the amino acid sequences of the invention.

Also, the invention relates to compounds that can modulate the (biological) activity of the amino acid sequences of the invention, to compositions that contain such compounds, and to the use of such compounds in the preparation of such compositions.

In particular, the invention relates to such compositions that are in the form of 5 pharmaceutical compositions, and more in particular in the form of pharmaceutical compositions for the prevention or treatment of metabolic diseases such as diabetes or obesity, and also to the use of compounds that can modulate the (biological) activity of the amino acid sequences of the invention in the preparation of such pharmaceutical compositions.

10 The invention further relates to the use of the nucleotide sequences, amino acid sequences, genetic constructs, host cells and/or host organisms of the invention in (methods for) identifying and/or developing compounds that can be used in the prevention or treatment of metabolic diseases.

15 Other aspects, embodiments, applications and advantages of the present invention will become clear from the further description below.

The present invention was established from the finding that the amino acid sequences of the invention are involved in metabolic processes (as further described below) and thus can be used as (potential) "target(s)" for *in vitro* and/or *in vivo* interaction with chemical compounds and other factors (with the term "target" having its 20 usual meaning in the art, vide for example the definition given in WO 98/06737), and also from the finding that the nucleic acid sequence and amino acid sequences of the invention are involved in metabolic diseases. Consequently, compounds and/or factors that have been identified as interacting with the amino acid sequences of the invention (e.g. by the methods as described hereinbelow) may be useful as active agents in the 25 pharmaceutical field, and in particular for the prevention and treatment of metabolic diseases. All this is supported by the following experimental data/observations:

- In an experimental model for fat handling, metabolism and storage using the model organism *C. elegans* (which model is further described in the Examples), downregulation of the *C. elegans* gene T17E9.1a by RNA interference strongly 30 reduces the fat storage phenotype in said nematode. As further described below, T17E9.1a is the *C. elegans* ortholog of STE20 like kinase in mammals.

Some particularly preferred examples of nucleotide sequences of the invention are:

- the nucleotide sequence of SEQ ID NO:1 (T17E9.1a, also known as *kin-18*), which is a sequence derived from the nematode worm *C. elegans*; and
- the human orthologs of said *C.elegans* sequence, as may be identified by

5 bioinformatic comparison of the *C.elegans* sequence with the human genome. Some preferred, but non-limiting orthologues are given in SEQ ID NOS: 3, 5, 7 and 9.

Generally herein, the use of the human nucleotide sequences of SEQ ID NOS: 3, 5, 7 and 9 and/or the use of nucleotide sequence derived therefrom will be preferred, in particular when the invention is used to develop compounds for pharmaceutical use.

10 In a broader sense, the term "*nucleotide sequence of the invention*" also comprises:

- parts and/or fragments of any of the nucleotide sequence of SEQ ID NO'S: 1, 3, 5, 7 and 9;
- (natural and/or synthetic) mutants, variants, alleles, analogs, orthologs (hereinbelow 15 collectively referred to as "*mutants*") of any of the nucleotide sequence of SEQ ID NO'S: 1-5, as further described below.
- parts and/or fragments of such (natural or synthetic) mutants;
- nucleotide fusions of any of the nucleotide sequence of SEQ ID NO'S:1-5 (or a part or fragment thereof) with at least one further nucleotide sequence;
- 20 - nucleotide fusions of (natural or synthetic) mutants (or a part or fragment thereof) with at least one further nucleotide sequence;

in which such mutants, parts, fragments and/or fusions are preferably as further described below.

25 The invention also comprises different splice variants of the above nucleotide sequences.

Some particularly preferred examples of amino acid sequences of the invention are:

- the amino acid sequence of SEQ ID NO:2, which is a sequence derived from the nematode worm *C. elegans*; and

- the human analogs of said *C.elegans* sequence, as may be identified by bioinformatic comparison of the *C.elegans* sequence with the human genome. Some preferred, but non-limiting analogs are given in SEQ ID NOS: 4, 6, 8 and 10.

In a broader sense, the term "*amino acid sequence of the invention*" also

5 comprises:

- parts and/or fragments of the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 and 10.
- (natural and/or synthetic) mutants, variants, alleles, analogs, orthologs (hereinbelow collectively referred to as "*analogs*") of the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 and 10;
- 10 - parts and/or fragments of such analogs;
- fusions of the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 and 10 (or a part or fragment thereof) with at least one further amino acid residue or sequence;
- fusions of the amino acid sequence of an analog (or a part or fragment thereof) with at least one further amino acid residue or sequence;
- 15 in which such mutants, parts, fragments and/or fusions are preferably as further described below.

The term "*amino acid sequence of the invention*" also comprises "immature" forms of the abovementioned amino acid sequences, such as a pre-, pro- or prepro-forms and/or fusions with suitable leader sequences. Also, the amino acid sequences of the 20 invention may have been subjected to post-translational processing and/or be suitably glycosylated, depending upon the host cell or host organism used to express/produce said amino acid sequence; and/or may be otherwise modified (e.g. by chemical techniques known per se in the art).

Generally herein, the use of the human amino acid sequence of SEQ ID NOS: 4, 25 6, 8 and 10 and/or of amino acid sequences derived therefrom will be preferred, in particular when the invention is used to develop compounds for pharmaceutical use.

From the prior art, the following is known about the nucleotide sequences, amino acid sequences and compounds of the invention:

(1) SEQ ID NOS: 1 and 2. These are the nucleotide sequence and amino acid sequence, respectively, of the *C. elegans* gene/protein T17E9.1a (also known as "*kin-18*")

30 Reference is made to the following entry from the WORMBASE database:

- for protein sequence:

<http://www.wormbase.org/db/seq/protein?name=WP%3ACE01405;class=Protein>

- for DNA sequence:

<http://www.wormbase.org/db/seq/sequence?name=T17E9.1a;class=Sequence>

5 as well as to Berman et al., Gene 279 (2001) 137-147.

(2) SEQ ID NOS: 3 and 4. These are the nucleotide sequence and amino acid sequence, respectively, of the human gene/protein known as "*JNK/SAPK inhibitory kinase*" (also known as "*JIK*", "*DPK*" or "*KDS*", or STE20-like kinase).

Reference is made to the following entries from the GENBANK database:

10

- NM_016281(mRNA); and
- NP_057365.2 (translation).

See also:

15

- GENBANK entry AF135158; Zhang et al., Biochem. Biophys. Res. Commun. 274 (3), 872-879 (2000); MEDLINE 20384190, PUBMED 10924369; and
- GENBANK entry AF179867; Tassi et al., J. Biol. Chem. 274 (47), 33287-33295 (1999); MEDLINE 20026851; PUBMED 10559204

(3) SEQ ID NOS: 5 and 6. These are the nucleotide sequence and amino acid sequence, respectively, of the human gene/protein known as "*prostate derived STE20-like kinase*" (also known as "*PSK*" or "*TAO2*").

20 Reference is made to the following entries from the GENBANK database:

- NM_016151.1 (mRNA)
- NP_057235.1 (translation)

See also:

25

- GENBANK entry AF061943; Moore et al, J. Biol. Chem. 275 (6), 4311-4322 (2000); MEDLINE 20127920; PUBMED 10660600;
- Chen et al., J. Biol. Chem., Vol. 274, No. 40, pp. 288803-28808 (1999).

(4) SEQ ID NOS: 7 and 8. These are the nucleotide sequence and amino acid sequence, respectively, of the human gene/protein known as "*thousand and one amino acid protein kinase*" (also known as "*TAO1*").

30 Reference is made to the following entries from the GENBANK database:

- NM_004783 (mRNA)

- NP_004774.1 (translation)

See also:

- GENBANK entry AB020688; Nagase et al, DNA Res. 5 (6), 355-364 (1998);
MEDLINE 99156230; PUBMED 10048485

5 (5) SEQ ID NOS: 9 and 10. These are the nucleotide sequence and amino acid sequence, respectively, of the human gene/protein which with designation KIAA1361 (note that although this gene and protein are also known as "*thousand and one amino acid protein kinase*" or "*TAO1*", KIAA1361 is a different gene/protein from SEQ ID NOS 7 and 8).

10 Reference is made to the following entries from the EMBL and SpTrEMBL databases

- AY049015 (EMBL)
- Q9P2I6 (SPTREMBL)

See also: Nagase et al., DNA Res. 7:65 (2000) and EMBL entry NR AB037782
(<http://www.ebi.ac.uk/cgi-bin/emblfetch?AB037782>)

15 Also, although the inventors do not wish to be limited thereto, it is believed that SEQ ID NOS: 5/6 on the one hand, and SEQ ID NOS: 7 and 8 on the other hand, are transcripts from the same gene/locus.

20 Thus, in a first aspect, the invention relates to a nucleic acid, preferably in (essentially) isolated form, which nucleic acid encodes and/or can be used to express an amino acid sequence of the invention (as defined herein), and in particular the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 or 10.

25 In another aspect, the invention relates to a nucleic acid, preferably in (essentially) isolated form, which nucleic acid comprises a nucleotide sequence of the invention, and in particular the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7 or 9.

30 In a yet another aspect, the invention relates to a nucleic acid, preferably in (essentially) isolated form, which nucleic acid essentially consists of a nucleotide sequence of the invention, and in particular a nucleotide sequence SEQ ID NO: 1, 3, 5, 7 or 9.

Collectively, these nucleic acids will also be referred to herein as "*nucleic acids of the invention*". Also, where appropriate in the context of the further description of the

invention below, the terms "*nucleotide sequence of the invention*" and "*nucleic acid of the invention*" may be considered essentially equivalent and essentially interchangeable.

Also, for the purposes of the present invention, a nucleic acid or amino acid sequence is considered to "*(in) essentially isolated (form)*" – for example, from its native biological source - when it has been separated from at least one other component (and in particular macromolecule) with which it is usually associated, such as another nucleic acid, another protein/polypeptide or another (polymeric) biological component. In particular, a nucleic acid or amino acid sequence is considered "*essentially isolated*" when it has been purified at least 2-fold, in particular at least 10-fold, more in particular at least 100-fold, and up to 1000-fold or more.

The nucleic acids of the invention may also be in the form of a genetic construct, again as further described below. These constructs will also be referred to herein as "*genetic constructs of the invention*". In a preferred embodiment, such a construct will comprise:

- 15 a) the nucleotide sequence of the invention; operably connected to:
- b) one or more regulatory elements, such as a promoter and optionally a suitable terminator;

and optionally also:

- c) one or more further elements of genetic constructs known per se;

20 in which the terms "*regulatory element*", "*promoter*", "*terminator*", "*further elements*" and "*operably connected*" have the meanings indicated hereinbelow.

In another aspect, the invention relates to a protein or polypeptide, preferably in (essentially) isolated form, said protein or polypeptide comprising an amino acid sequence of the invention (as defined above), and in particular the amino acid sequence 25 of SEQ ID NOS: 2, 4, 6, 8 or 10.

In a further aspect, the invention relates to a protein or polypeptide, preferably in (essentially) isolated form, said protein or polypeptide essentially consisting of an amino acid sequence of the invention (as defined above), and in particular of the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 or 10.

30 In a further aspect, the invention relates to methods for transforming a host cell and/or a host organism with a nucleotide sequence, with a nucleic acid and/or with a

genetic construct of the invention. The invention also relates to the use of a nucleotide sequence, of a nucleic acid and/or of a genetic construct of the invention transforming a host cell or a host organism.

In yet another aspect, the invention relates to a host cell or host organism that has
5 been transformed and/or contains with a nucleotide sequence, with a nucleic acid and/or with a genetic construct of the invention. The invention also relates to a host cell and/or host organism that expresses, or (at least) is capable of expressing (e.g. under suitable conditions), an amino acid sequence of the invention. Collectively, such host cells/host organisms will also be referred to herein as "*host cells/host organisms of the invention*".

10 In yet another aspect, the invention relates to methods for producing an amino acid sequence of the invention, in which a nucleotide sequence, nucleic acid, genetic construct, host cell or host organism of the invention is used. Such methods may for instance include expressing a nucleotide sequence of the invention in a suitable host cell or host organism (e.g. upon suitable transformation), and/or maintaining and/or
15 cultivating a host cell or host organism of the invention under suitable conditions, i.e. such that an amino acid sequence of the invention is expressed or obtained. Optionally, these methods may also comprise (one or more steps for) isolating the amino acid sequence thus expressed/produced. The invention also relates to the use of a nucleotide sequence, a nucleic acid, a genetic construct and/or a host cell/host organism of the
20 invention in such a method.

25 In yet a further aspect, the invention relates to a method for identifying a compound that can modulate the (biological) activity of, and/or that can otherwise interact with, an amino acid sequence of the invention, which method is as further described below. The invention also relates to the use of a nucleotide sequence, a nucleic acid, a genetic construct, an amino acid sequence and/or a host cell/host organism of the invention in such a method.

30 In yet a further aspect, the invention relates to a method for identifying a compound that can be used in (the preparation of a pharmaceutical composition for) the prevention and/or treatment of metabolic diseases (as further defined below), which method is as further described below. The invention also relates to the use of a nucleotide

sequence, a nucleic acid, a genetic construct, an amino acid sequence and/or a host cell/host organism of the invention in such a method.

The invention also relates to compounds that can modulate the (biological activity of), and/or that can otherwise interact with, an amino acid sequence of the invention, 5 either *in vitro* or preferably (also) *in vivo*, as further described below. The invention also relates to compositions that contain such compounds, and in particular to pharmaceutical compositions that contains such compounds.

The invention further relates to the use of compounds that can modulate the (biological activity of), and/or that can otherwise interact with, an amino acid sequence of 10 the invention in the preparation of these compositions, and in particular to the use of such compounds in the preparation of a pharmaceutical composition for the prevention and/or treatment of metabolic diseases.

The invention also relates to compounds that can be used in the prevention and/or treatment of metabolic diseases (as further defined below), which compounds have or can 15 be identified and/or developed using the method, nucleic acid sequence, amino acid sequence and/or host cell or host organism of the invention. The invention also relates to compositions that contain such compounds, and in particular to pharmaceutical compositions that contain said compounds.

The invention also relates to the use of such compounds in the preparation of a 20 pharmaceutical composition, and in particular to the use of such compounds in the preparation of a pharmaceutical composition for the prevention or treatment of metabolic diseases.

Unless explicitly specified herein, all terms used in the present description have 25 their usual meaning in the art, for which particular reference is made to the definitions given in WO98/06737 and EP 1 085 089.

The nucleotide sequences and amino acid sequences of the invention may generally further be characterized by the presence of a kinase domain (identified using SMART™-analysis), as follows:

30 - SEQ ID NO:2: from amino acid (a.a.) residue 30 to a.a. residue 289

- SEQ ID NO:4: from a.a. residue 24 to a.a. residue 277
- SEQ ID NO:6: from a.a. residue 28 to a.a. residue 281
- SEQ ID NO:8: from a.a. residue 28 to a.a. residue 281
- 5 - SEQ ID NO:10: from a.a. residue 32 to a.a. residue 285

On the basis of the above and on the basis of the prior art referred to above, and although the invention is not specifically limited to any specific explanation or mechanism, it is assumed that the nucleotide sequences and/or amino acid sequences
10 have (biological) activity as kinases.

In particular, and although the invention is again not limited to any specific explanation or hypothesis, it is assumed that the amino acid sequences of the invention may be involved in (the modulation of) the JNK/SAPK pathway (vide for example Zhang et al. and Tassi et al., both *supra*; as well as Yoneda et al., *J. Biol. Chem.*, Vol. 276, 15 No.17, p. 13935-13940 (2001)).

For JNKS and their role in diabetes/obesity, reference is *inter alia* made to Hirosumi et al., *Nature*, Vol. 20, November 2002, 333-336.

As is known in the art, biological activity of this kind can be measured using standard assay techniques for kinases, which are well known to the skilled person. Some
20 preferred, but non-limiting examples include:

- the JNK/SAPK assay described by Zhang et al., *supra*, which is particularly suitable for the kinase of SEQ ID NO 4 or a similar kinase;
- the kinase assays described by Moore et al., *supra*, in particular the *in vitro* kinase assays described on page 4315;

25 The nucleotide sequence of SEQ ID NO: 1 was identified, and can be derived/isolated from/using the nematode *C.elegans*; in the manner as further described in Berman et al., or in any other suitable manner known *per se*.

The nucleotide sequences of SEQ ID NOS: 3, 5, 7 and 9 were identified, and can be derived/isolated from/using human cells; in the manner as further described in the
30 prior art referred to above, or in any other suitable manner known *per se*.

Also, it is expected that - based upon the disclosure herein - the skilled person will be able to identify, derive and/or isolate natural "mutants" (as mentioned above) of the nucleotide sequence of SEQ ID NOS: 1, 3 , 5, 7 and/or 9. For example, such mutants could be derived from (other individuals of) the same species (for example from an individual of a different strain or line, including but not limited to mutant strains or lines); and/or from (individuals of) other species (in which case these mutants will also be referred to herein as "*orthologs*"). Some examples of species from which such orthologs could be derived include, but are not limited to species of

- unicellular and/or micro-organisms such as bacteria, and yeast,
- invertebrate multicellular organisms as such as insects and nematodes (for example, agronomically harmful insect or nematode species);
- vertebrate multicellular organisms as such as fish, birds, reptiles, amphibians and mammals;

Preferably, a natural ortholog is derived from a mammal such as a mouse, rat, rabbit or dog.

Such natural mutants may be obtained in a manner essentially analogous to the methods described in the prior art referred to above, or alternatively by:

- construction of a DNA library from the species of interest in an appropriate expression vector system, followed by direct expression of the mutant sequence;
- construction of a DNA library from the species of interest in an appropriate expression vector system, followed by screening of said library with a probe of the invention (as described below) and/or with a(nother) nucleotide sequence of the invention;
- isolation of mRNA that encodes the mutant sequence from the species of interest, followed by cDNA synthesis using reverse transcriptase;

and/or by any other suitable method(s) or technique(s) known per se, for which reference is for instance made to the standard handbooks, such as Sambrook et al, "Molecular Cloning: A Laboratory Manual" (2nd.ed.), Vols. 1-3, Cold Spring Harbor Laboratory Press (1989) and F. Ausubel et al, eds., "Current protocols in molecular biology", Green Publishing and Wiley Interscience, New York (1987).

It is also expected that - based upon the disclosure herein - the skilled person will be able to provide and/or derive synthetic mutants (as defined hereinabove) of the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7 and/or 9.

Techniques for generating such synthetic sequences will be clear to the skilled 5 person and may for instance include, but are not limited to, automated DNA synthesis; site-directed mutagenesis; combining two or more parts of one or more naturally occurring sequences, introduction of mutations that lead to the expression of a truncated expression product; introduction of one or more restriction sites (e.g. to create cassettes and/or regions that may easily be digested and/or ligated using suitable restriction 10 enzymes), and/or the introduction of mutations by means of a PCR reaction using one or more "mismatched" primers, using for example a sequence of a naturally occurring GPCR as a template. These and other techniques will be clear to the skilled person, and reference is again made to the standard handbooks, such as Sambrook et al. and Ausubel et al., mentioned above.

15 Preferably, any mutants as described herein will encode amino acid sequences having one or more, and preferably all, of the structural characteristics/conserved features referred to above for the sequences of SEQ ID NO: 2, 4, 6, 8 and/or 10, and in particular will contain a kinase domain.

It is also possible in the invention to use a part or fragment of the nucleotide 20 sequences of SEQ ID NOS 1, 3, 5, 7 or 9; or a part or fragment of a (natural or synthetic) mutant thereof. These may for instance be 5' and/or 3' truncated nucleotide sequences, or sequences with an introduced *in frame* startcodon or stopcodon. Also, two or more such parts or fragments of one or more nucleotide sequences of the invention may be suitably combined (e.g. ligated in frame) to provide a (further) nucleotide sequence of the 25 invention.

Preferably, any such parts or fragments will be such that they comprise at least one continuous stretch of at least 15 nucleotides, preferably at least 30 nucleotides, more preferably at least 60 nucleotides, even more preferably more than 90 nucleotides, of one or more of the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7 and/or 9.

30 In particular, any mutants, parts or fragments as described herein may be such that they (at least) encode the active/catalytic site of the corresponding amino acid sequence

of the invention and/or a binding domain of the corresponding amino acid sequence of the invention

Any mutants, parts and/or fragments as described herein are preferably (also) such that they are capable of hybridizing with one or more of the nucleotide sequences of SEQ 5 ID NOS 1, 3, 5, 7 and/or 9, i.e. under conditions of "moderate stringency", and preferably under conditions of "high stringency". Such conditions will be clear to the skilled person, for example from the standard handbooks, such as Sambrook et al. and Ausubel et al., mentioned above, as well as in EP 0 967 284, EP 1 085 089 or WO 00/55318.

In particular, any mutants, parts and/or fragments as described herein may be such 10 that they are capable of hybridizing with the nucleotide sequence of SEQ ID NO 1, 3, 5, 7 and/or 9 under the "stringent" hybridisation conditions described in WO 00/78972 and WO 98/49185, and/or under the hybridization conditions described in GB 2 357 768-A.

Also, any mutants, parts and/or fragments as described herein will preferably have 15 a degree of "sequence identity", at the nucleotide level, with one or more of the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7 and/or 9 of at least 50%, preferably at least 60%, more preferably at least 70%, even more preferably at least 80%, and in particular more than 90%, and up to 95% or more.

For this purpose, the percentage of "sequence identity" between a given 20 nucleotide sequence and the nucleotide sequence of SEQ ID NO:1, 3, 5, 7 and/or 9 may be calculated by dividing [*the number of nucleotides in the given nucleotide sequence that are identical to the nucleotide at the corresponding position in the nucleotide sequence of the relevant SEQ ID NO*] by [*the total number of nucleotides in the given nucleotide sequence*] and multiplying by [100%], in which each deletion, insertion, 25 substitution or addition of a nucleotide - compared to the sequence of the relevant SEQ ID NO - is considered as a difference at a single nucleotide (position).

Alternatively, the degree of sequence identity may be calculated using a known computer algorithm for sequence alignment such as NCBI Blast v2.0, using standard settings.

Some other techniques, computer algorithms and settings for determining the 30 degree of sequence identity are for example described in EP 0 967 284, EP 1 085 089, WO 00/55318, WO 00/78972, WO 98/49185 and GB 2 357 768-A.

Also, in a preferred aspect, any mutants, parts and/or fragments as described herein will encode proteins/polypeptides having a biological activity that is essentially similar to the biological activity described above for the sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10, i.e. to a degree of at least 10%, preferably at least 50 % more preferably at 5 least 75%, and up to 90%, as measured by a suitable assay for kinase activity known per se, such as those mentioned above (preferred) or below. In particular, they will have activity as a kinase, as measured using a suitable assay of kinase activity, such as those referred to above, and preferably an activity which is at least 10%, preferably at least 50 % more preferably at least 75%, and up to 90% or more, of the kinase activity of any one 10 of the sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10.

Preferably, any mutants, parts and/or fragments of the nucleotide sequence of the invention will (also) be such that they encode an amino acid sequence which has a degree of "sequence identity", at the amino acid level, with one or more of the amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 and/or 10 of at least 50%, preferably at least 60%, 15 more preferably at least 70%, even more preferably at least 80%, and in particular more than 90% and up to 95 % or more, in which the percentage of "sequence identity" is calculated as described below.

Preferably, a nucleotide sequence of the invention will (also) have a length (expressed as total number of nucleotides), which is at least 50%, preferably at least 60%, 20 more preferably at least 70%, even more preferably at least 80%, and in particular more than 90% and up to 95 % or more of the length of one or more of the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7 and/or 9.

Generally, the nucleotide sequences of the invention, when in the form of a nucleic acid, may be DNA or RNA, and may be single stranded or double stranded. For 25 example, the nucleotide sequences of the invention may be genomic DNA, cDNA or synthetic DNA (such as DNA with a codon usage that has been specifically adapted for expression in the intended host cell or host organism). Thus, the nucleotide sequences of the invention may contain intron sequences, and also generally comprises different splice variants.

30 It is also within the scope of the invention to use a fusion of a nucleotide sequence of the invention (as described above) with one or more further nucleotide sequence(s),

including but not limited to one or more coding sequences, non-coding sequences and/or regulatory sequences. Preferably, in such fusions, the one or more further nucleotide sequences are operably connected (as described below) to the nucleotide sequence of the invention (for example so that, when the further nucleotide sequence is a coding sequence, the nucleotide fusion encodes a protein fusion as described below).

Another embodiment of the invention relates to a nucleic acid probe that is capable of hybridizing with a nucleotide sequence of the invention under conditions of moderate stringency, preferably under conditions of high stringency, and in particular under stringent conditions (all as described above). Such nucleotide probes may for 10 instance be used for detecting and/or isolating a(nother) nucleotide sequence of the invention and/or as a primer for amplifying a nucleotide sequence of the invention; all using techniques known *per se*, for which reference is again made to the general handbooks such as Sambrook et al. and Ausubel et al. mentioned above.

Generally, such probes can be designed by the skilled person starting from a 15 nucleotide sequence and/or amino acid sequence of the invention - and in particular one or more of the sequences of SEQ ID NOS 1, 3, 5, 7 and/or 9 - optionally using a suitable computer algorithm. Also, as will be clear to the skilled person, such probes may be degenerate probes.

In another embodiment, the invention relates to an antisense molecule against a 20 nucleotide sequence of the invention.

Yet another embodiment relates to a double stranded RNA molecule directed against a nucleotide sequence of the invention (one strand of which will usually comprise at least part of a nucleotide sequence of the invention). The invention also relates to 25 genetic constructs that can be used to provide such double stranded RNA molecules (e.g. by suitable expression in a host cell or host organism, or for example in a bacterial strain such as *E.coli*). For such constructs, reference is made to for example the International Applications PCT/IB01/1068 and WO 00/01846, both by applicant.

The amino acid sequence of SEQ ID NO: 2 was identified, and can be 30 derived/isolated from/using the nematode *C.elegans*; in the manner as further described in Berman et al., or in any other suitable manner known *per se*.

The amino acid sequences of SEQ ID NOS: 4, 6, 8 and 10 were identified, and can be derived/isolated from/using human cells; in the manner as further described in the prior art referred to above, or in any other suitable manner known per se.

Generally, the amino acid sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10 may be 5 isolated from the species mentioned above (i.e. *C. elegans* and *human*, respectively), using any technique(s) for protein isolation and/or purification known per se. Alternatively, the amino acid sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10 may be obtained by suitable expression of a suitable nucleotide sequence - such as one of the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7 and/or 9, as applicable or a suitable 10 mutant thereof - in an appropriate host cell or host organism, as further described below.

Also, it is expected that - based upon the disclosure herein - the skilled person will be able to identify, derive and/or isolate natural "analogs" (as mentioned above) of the amino acid sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10. Such mutants could be derived from (other individuals of) the same species (for example from an individual of a 15 different strain or line, including but not limited to mutant strains or lines); and/or from (individuals of) other species. For example, such analogs could be derived from the insect species or other pest species mentioned above.

Such natural analogs may again be obtained by isolating them from their natural source using any technique(s) for protein isolation and/or purification known per se, or 20 alternatively by suitable expression of a suitable nucleotide sequence of the invention - such as a natural mutant as described above - in an appropriate host cell or host organism, as further described below.

It is also expected that - based upon the disclosure herein - the skilled person will be able to provide and/or derive synthetic "analogs" (as mentioned above) of one or 25 more of the amino sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10.

Generally, such synthetic analogs may be obtained by suitable expression of a suitable nucleotide sequence of the invention - such as a synthetic mutant as described above - in an appropriate host cell or host organism, as further described below.

Preferably, any analogs as described herein will have one or more, and preferably 30 all, of the structural characteristics/conserved features referred to above for the sequences of SEQ ID NO: 2, 4, 6, 8 and/or 10, and in particular will contain a kinase domain.

It is also possible in the invention to use a part or fragment of one or more of the amino acid sequences of SEQ ID NOS 2, 4, 6, 8 and/or 10, or a part or fragment of a (natural or synthetic) analog thereof mutant thereof. This may for instance be N- and/or C- truncated amino acid sequence. Also, two or more parts or fragments of one or more 5 amino acid sequences of the invention may be suitably combined to provide a (further) amino acid sequence of the invention.

Preferably, any such parts or fragments will be such that they comprise at least one continuous stretch of at least 5 amino acids, preferably at least 10 amino acids, more preferably at least 20 amino acids, even more preferably more than 30 amino acids, of 10 one or more of the amino acid sequences of SEQ ID NO: 2, 4, 6, 8 and/or 10.

In particular, any parts or fragments as described herein are such that they (at least) comprise the active/catalytic site of the corresponding amino acid sequence of the invention and/or a binding domain of the corresponding amino acid sequence of the invention, and in particular a kinase domain. As will be clear to the skilled person, such 15 parts or fragments may find particular use in assay- and screening techniques (as generally described below) and/or (when said part or fragment is provided in crystalline form) in X-ray crystallography.

Generally, such parts or fragments of the amino acid sequences of the invention may be obtained by suitable expression of a suitable nucleotide sequence of the invention 20 - such as a suitable part or fragment as described hereinabove for the nucleotide sequences of the invention - in an appropriate host cell or host organism, as further described below.

In addition and/or as an alternative to the methodology above, amino acid sequences of the invention may also be provided by (chemically and/or enzymatically) 25 modifying the side chain(s) of one or more amino acid residues of an amino acid sequence of SEQ ID NO: 2, 4, 6, 8 and/or 10 or a part, fragment, (natural and/or synthetic) mutant, variant, allele, analogs, orthologs thereof, for example by one or more of the side chain modifications as described in WO 01/02560 and/or by incorporating (e.g. by insertion and/or substitution) one or more unnatural amino acid residues, again as 30 described in WO 01/02560.

Preferably, any analogs, parts and/or fragments as described herein will be such that they have a degree of "sequence identity", at the amino acid level, with one or more of the amino acid sequences of SEQ ID NOS 2, 4, 6, 8 and/or 10 of at least 50%, preferably at least 60%, more preferably at least 70%, even more preferably at least 80%,
5 and in particular more than 90% and up to 95 % or more.

For this purpose, the percentage of "sequence identity" between a given amino acid sequence and one of the amino acid sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10 may be calculated by dividing [*the number of amino acid residues in the given amino acid sequence that are identical to the amino acid residue at the corresponding position in the amino acid sequence of the relevant SEQ ID NO*] by [*the total number of amino acid residues in the given amino acid sequence*] and multiplying by [100%], in which each deletion, insertion, substitution or addition of an amino acid residue - compared to the sequence of the relevant SEQ ID NO - is considered as a difference at a single amino acid (position).
10

15 Alternatively, the degree of sequence identity may be calculated using a known computer algorithm, such as those mentioned above.

Also, such sequence identity at the amino acid level may take into account so-called "conservative amino acid substitutions", which are well known in the art, for example from GB-A-2 357 768, WO 98/49185, WO 00/46383 and WO 01/09300; and
20 (preferred) types and/or combinations of such substitutions may be selected on the basis of the pertinent teachings from the references mentioned in WO 98/49185.

Also, preferably, any analogs, parts and/or fragments as described herein will have a biological activity that is essentially similar to the biological activity described above for the sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10, i.e. to a degree of at least
25 10%, preferably at least 50 %, more preferably at least 75%, and up to 90%, as measured by the assay mentioned above. In particular, they will have activity as a kinase, as measured using a suitable assay of kinase activity, such as those referred to above, and preferably an activity which is at least 10%, preferably at least 50 %, more preferably at least 75%, and up to 90% or more, of the kinase activity of any one of the sequences of
30 SEQ ID NOS: 2, 4, 6, 8 and/or 10.

Preferably, an amino acid sequence of the invention will (also) have a length (expressed as total number of amino acid residues), which is at least 50%, preferably at least 60%, more preferably at least 70%, even more preferably at least 80%, and in particular more than 90% and up to 95 % or more of the length of one or more of the 5 amino acid sequence of SEQ ID NOS: 2, 4, 6, 8 and/or 10.

It is also within the scope of the invention to use a fusion of an amino acid sequence of the invention (as described above) with one or more further amino acid sequences, for example to provide a protein fusion. Generally, such fusions may be obtained by suitable expression of a suitable nucleotide sequence of the invention - such 10 as a suitable fusion of a nucleotide sequence of the invention with one or more further coding sequences - in an appropriate host cell or host organism, as further described below. According to one non-limiting embodiment, such a fusion comprises at least the kinase domain of one of the amino acid sequences of the invention.

In one particular embodiment, such fusions may comprise an amino acid sequence 15 of the invention (that preferably at least comprises the kinase domain, and may essentially consist of the kinase domain) fused with a reporter protein such as GFP, luciferase or another fluorescent protein moiety. As will be clear to the skilled person, such fusions may find particular use in expression analysis and similar methodologies.

In another embodiment, the fusion partner may be an amino acid sequence or 20 residue that may be used in purification of the expressed amino acid sequence, for example using affinity techniques directed against said sequence or residue. Thereafter, said sequence or residue may be removed (e.g. by chemical or enzymatical cleavage) to provide the nucleotide sequence of the invention (for this purpose, the sequence or residue may optionally be linked to the amino acid sequence of the invention via a 25 cleavable linker sequence). Some preferred, but non-limiting examples of such residues are multiple histidine residues and glutatione residues,

In one preferred, but non-limiting aspect, any such fusion will have a biological 30 activity that is essentially similar to the biological activity described above for the sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10, i.e. to a degree of at least 10%, preferably at least 50 %, more preferably at least 75%, and up to 90%, as measured by the assay mentioned above. In particular, they will have activity as a kinase, as measured

using a suitable assay of kinase activity, such as those referred to above, and preferably an activity which is at least 10%, preferably at least 50 %, more preferably at least 75%, and up to 90% or more, of the kinase activity of any one of the sequences of SEQ ID NOS: 2, 4, 6, 8 and/or 10.

5 Genetic constructs of the invention will generally comprise at least one nucleotide sequence of the invention, optionally linked to one or more elements of genetic constructs known *per se*, as described below.

Such genetic constructs may be DNA or RNA, and are preferably double-stranded DNA. The constructs may also be in a form suitable for transformation of the intended 10 host cell or host organism, in a form suitable for integration into the genomic DNA of the intended host cell or in a form suitable for independent replication, maintenance and/or inheritance in the intended host organism. For instance, the genetic construct may be in the form of a vector, such as for example a plasmid, cosmid, YAC, a viral vector or transposon. In particular, the vector may be an expression vector, i.e. a vector that can 15 provide for expression *in vitro* and/or *in vivo* (e.g. in a suitable host cell and/or host organism as described below).

As the one or more "further elements" referred to above, the genetic construct(s) of the invention may generally contain one or more suitable regulatory elements (such as 20 a suitable promoter(s), enhancer(s), terminator(s), etc.), 3'- or 5'-UTR sequences, leader sequences, selection markers, expression markers/reporter genes, and/or elements that may facilitate or increase (the efficiency of) transformation or integration. These and other suitable elements for such genetic constructs will be clear to the skilled person, and may for instance depend upon the type of construct used, the intended host cell or host organism; the manner in which the nucleotide sequences of the invention of interest are to 25 be expressed (e.g. via constitutive, transient or inducible expression); and/or the transformation technique to be used.

Preferably, in the genetic constructs of the invention, the one or more further elements are "*operably linked*" to the nucleotide sequence(s) of the invention and/or to 30 each other, by which is generally meant that they are in a functional relationship with each other. For instance, a promoter is considered "*operably linked*" to a coding sequence if said promoter is able to initiate or otherwise control/regulate the transcription and/or

the expression of a coding sequence (in which said coding sequence should be understood as being "*under the control of*" said promotor)

Generally, when two nucleotide sequences are operably linked, they will be in the same orientation and usually also in the same reading frame. They will usually also be

5 essentially contiguous, although this may also not be required.

Preferably, the optional further elements of the genetic construct(s) used in the invention are such that they are capable of providing their intended biological function in the intended host cell or host organism.

For instance, a promoter, enhancer or terminator should be "*operable*" in the

10 intended host cell or host organism, by which is meant that (for example) said promoter should be capable of initiating or otherwise controlling/regulating the transcription and/or the expression of a nucleotide sequence - e.g. a coding sequence - to which it is operably linked (as defined above).

Such a promoter may be a constitutive promoter or an inducible promoter, and

15 may also be such that it (only) provides for expression in a specific stage of development of the host cell or host organism, and/or such that it (only) provides for expression in a specific cell, tissue, organ or part of a multicellular host organism.

Some particularly preferred promoters include, but are not limited to those present in the expression vectors referred to below.

20 A selection marker should be such that it allows - i.e. under appropriate selection conditions - host cells and/or host organisms that have been (successfully) transformed with the nucleotide sequence of the invention to be distinguished from host cells/organisms that have not been (successfully) transformed. Some preferred, but non-limiting examples of such markers are genes that provide resistance against antibiotics

25 (such as kanamycine or ampicilline), genes that provide for temperature resistance, or genes that allow the host cell or host organism to be maintained in the absence of certain factors, compounds and/or (food) components in the medium that are essential for survival of the non-transformed cells or organisms.

30 A leader sequence should be such that - in the intended host cell or host organism - it allows for the desired post-translational modifications and/or such that it directs the transcribed mRNA to a desired part or organelle of a cell. A leader sequence may also

allow for secretion of the expression product from said cell. As such, the leader sequence may be any pro-, pre-, or prepro-sequence operable in the host cell or host organism.

An expression marker or reporter gene should be such that - in the host cell or host organism - it allows for detection of the expression of (a gene or nucleotide sequence 5 present on) the genetic construct. An expression marker may optionally also allow for the localisation of the expressed product, e.g. in a specific part or organelle of a cell and/or in (a) specific cell(s), tissue(s), organ(s) or part(s) of a multicellular organism. Such reporter genes may also be expressed as a protein fusion with the amino acid sequence of the invention. Some preferred, but non-limiting examples include fluorescent proteins such 10 as Green Fluorescent Protein (GFP).

For some (further) non-limiting examples of the promoters, selection markers, leader sequences, expression markers and further elements that may be present/used in the genetic constructs of the invention - such as terminators, transcriptional and/or translational enhancers and/or integration factors - reference is made to the general 15 handbooks such as Sambrook et al. and Ausubel et al. mentioned above, to W.B. Wood et al., "The nematode *Caenorhabditis elegans*", Cold Spring Harbor Laboratory Press (1988) and D.L. Riddle et al., "C. ELEGANS II", Cold Spring Harbor Laboratory Press (1997), as well as to the examples that are given in WO 95/07463, WO 96/23810, WO 95/07463, WO 95/21191, WO 97/11094, WO 97/42320, WO 98/06737, WO 98/21355, 20 US-A-6,207,410, US-A- 5,693,492 and EP 1 085 089. Other examples will be clear to the skilled person.

The genetic constructs of the invention may generally be provided by suitably linking the nucleotide sequence(s) of the invention to the one or more further elements described above, for example using the techniques described in the general handbooks 25 such as Sambrook et al. and Ausubel et al., mentioned above.

Often, the genetic constructs of the invention will be obtained by inserting a nucleotide sequence of the invention in a suitable (expression) vector known per se. Some preferred, but non-limiting examples of suitable expression vectors include:

- vectors for expression in mammalian cells: pMAMneo (Clontech), pcDNA3 30 (Invitrogen), pMC1neo (Stratagene), pSG5 (Stratagene), EBO-pSV2-neo (ATCC 37593), pBPV-1 (8-2) (ATCC 37110), pdBPV-MMTneo (342-12) (ATCC 37224),

pRSVgpt (ATCC37199), pRSVneo (ATCC37198), pSV2-dhfr (ATCC 37146), pUCTag (ATCC 37460) and 1ZD35 (ATCC 37565);

- vectors for expression in bacterial cells: pET vectors (Novagen) and pQE vectors (Qiagen);
- 5 - vectors for expression in yeast or other fungal cells: pYES2 (Invitrogen) and Pichia expression vectors (Invitrogen);
- vectors for expression in insect cells: pBlueBacII (Invitrogen).

The nucleotide sequences and/or genetic constructs of the invention may be used to transform a host cell or host organism.

10 The host cell may be any suitable (fungal, prokaryotic or eukaryotic) cell or cell line, for example:

- a bacterial strain, including but not limited to strains of *E.coli*, *Bacillus*., *Streptomyces* and *Pseudomonas*;
- a fungal cell, including but not limited to cells from species of *Aspergillus* and 15 *Trichoderma*;
- a yeast cell, including but not limited to cells from species of *Kluyveromyces* or *Saccharomyces*;
- an amphibian cell or cell line, such as *Xenopus* oocytes.

20 In one specific embodiment, which may prove particularly useful when the nucleotide sequences of the invention are (to be) used in the discovery and development of insecticidal compounds, the host cell may be an insect-derived cell or cell line, such as:

- cells/cell lines derived from *lepidoptera*, including but not limited to *Spodoptera* SF9 and Sf21 cells,
- 25 - cells/cell lines derived from *Drosophila*, such as Schneider and Kc cells; and/or
- cells/cell lines derived from a pest species of interest (as mentioned below), such as from *Heliothis virescens*.

In one preferred embodiment, the host cell is a mammalian cell or cell line, for example derived from the mammals referred to above.

In an even more preferred aspect, the host cell is a cell or cell line derived from a human, and from other mammals including but not limited to CHO- and BHK-cells and human cells or cell lines such as HeLa and COS.

In one specific, but non-limiting embodiment, the cell or cell line may be a human cell or cell line which is related to metabolic processes or metabolic disease and/or used as a cellular model for metabolic disease, including but not limited to liver cells or cell lines, adipocytes or muscle cells or cell lines such as HEPG2 cells, 3T3L1 adipocytes, CTC12 cells and L6 myotubes.

The host organism may be any suitable multicellular (vertebrate or invertebrate) 10 organism, including but not limited to:

- a nematode, including but not limited to nematodes from the genus *Caenorhabditis*, such as *C.elegans*,
- an insect, including but not limited to species of *Drosophila* and/or a specific pest species of interest (such as those mentioned above);
- other well known model organisms, such as zebrafish;
- a mammal such as a rat or mouse;

Other suitable host cells or host organisms will be clear to the skilled person, for example from the handbooks and patent applications mentioned above.

It should be noted that when a nucleotide sequence of the invention is expressed 20 in a multicellular organism, it may be expressed throughout the entire organism, or only in one or more specific cells, tissues, organs and/or parts thereof, for example by expression under the control of a promoter that is specific for said cell(s), tissue(s), organ(s) or part(s).

The nucleotide sequence may also be expressed during only a specific stage of 25 development or life cycle of the host cell or host organism, again for example by expression under the control of a promoter that is specific for said stage of development or life cycle. Also, as already mentioned above, said expression may be constitutive, transient and/or inducible.

According to one specific embodiment, the expression of a nucleotide sequence of 30 the invention in a host cell or host organism may be partly or totally reduced (i.e. knocked out), compared to the original (e.g. native) host cell or host organism. This may for

instance be achieved in a transient manner using antisense and/or RNA-interference techniques well known in the art, or in a constitutive manner using random, site specific and/or chemical mutagenesis of the nucleotide sequence of the invention, or any other suitable techniques for generating "knock-down" or "knock-out" animals.

5 Suitable transformation techniques will be clear to the skilled person and may depend on the intended host cell/host organism and the genetic construct to be used. Some preferred, but non-limiting examples of suitable techniques include ballistic transformation, (micro-)injection, transfection (e.g. using suitable transposons), electroporation and lipofection. For these and other suitable techniques, reference is again 10 made to the handbooks and patent applications mentioned above.

After transformation, a step for detecting and selecting those host cells or host organisms that have been successfully transformed with the nucleotide sequence/genetic construct of the invention may be performed. This may for instance be a selection step based on a selectable marker present in the genetic construct of the invention or a step 15 involving the detection of the amino acid sequence of the invention, e.g. using specific antibodies.

The transformed host cell (which may be in the form of a stable cell line) or host organisms (which may be in the form of a stable mutant line or strain) form further aspects of the present invention.

20 Preferably, these host cells or host organisms are such that they express, or are (at least) capable of expressing (e.g. under suitable conditions), an amino acid sequence of the invention (and in case of a host organism: in at least one cell, part, tissue or organ thereof). The invention also includes further generations, progeny and/or offspring of the host cell or host organism of the invention, that may for instance be obtained by cell 25 division or by sexual or asexual reproduction.

To produce/obtain expression of the amino acid sequences of the invention, the transformed host cell or transformed host organism may generally be kept, maintained and/or cultured under conditions such that the (desired) amino acid sequence of the invention is expressed-produced. Suitable conditions will be clear to the skilled person 30 and will usually depend upon the host cell/host organism used, as well as on the regulatory elements that control the expression of the (relevant) nucleotide sequence of

the invention. Again, reference is made to the handbooks and patent applications mentioned above in the paragraphs on the genetic constructs of the invention.

Generally, suitable conditions may include the use of a suitable medium, the presence of a suitable source of food and/or suitable nutrients, the use of a suitable 5 temperature, and optionally the presence of a suitable inducing factor or compound (e.g. when the nucleotide sequences of the invention are under the control of an inducible promoter); all of which may be selected by the skilled person. Again, under such conditions, the amino acid sequences of the invention may be expressed in a constitutive manner, in a transient manner, or only when suitably induced.

10 It will also be clear to the skilled person that the amino acid sequence of the invention may (first) be generated in an immature form (as mentioned above), which may then be subjected to post-translational modification, depending on the host cell/host organism used. Also, the amino acid sequence of the invention may be glycosylated, again depending on the host cell/host organism used.

15 The amino acid sequence of the invention may then be isolated from the host cell/host organism and/or from the medium in which said host cell or host organism was cultivated, using protein isolation and/or purification techniques known per se, such as (preparative) chromatography and/or electrophoresis techniques, differential precipitation techniques, affinity techniques (e.g. using a specific, cleavable amino acid sequence 20 fused with the amino acid sequence of the invention) and/or preparative immunological techniques (i.e. using antibodies against the amino acid sequence to be isolated).

In one embodiment, the amino acid sequence thus obtained may also be used to generate antibodies specifically against said sequence or an antigenic part or epitope thereof.

25 Such antibodies, which form a further aspect of the invention, may be generated in a manner known per se, for example as described in GB-A-2 357 768, US-A-5,693,492, WO 95/32734, WO 96/23882, WO 98/02456, WO 98/41633 and/or WO 98/49306, and/or as described in the prior art referred to above. Often, but not exclusively, such methods will involve immunizing an immunocompetent host with the pertinent 30 amino acid sequence of the invention or an immungenic part thereof (such as a specific epitope), in amount(s) and according to a regimen such that antibodies against said amino

acid sequence are raised, and than harvesting the antibodies thus generated, e.g. from blood or serum derived from said host.

For instance, polyclonal antibodies can be obtained by immunizing a suitable host such as a goat, rabbit, sheep, rat, pig or mouse with (an epitope of) an amino acid sequence 5 of the invention, optionally with the use of an immunogenic carrier (such as bovine serum albumin or keyhole limpet hemocyanin) and/or an adjuvant such as Freund's, saponin, ISCOM's, aluminium hydroxide or a similar mineral gel, or keyhole limpet hemocyanin or a similar surface active substance. After a suitable immune response has been raised (usually 10 within 1-7 days), the antibodies can be isolated from blood or serum taken from the immunized animal in a manner known per se, which optionally may involve a step of screening for an antibody with desired properties (i.e. specificity) using known immunoassay techniques, for which reference is again made to for instance WO 96/23882.

Monoclonal antibodies may, for example, be produced using continuous cell lines in culture, including hybridoma-based and similar techniques, again essentially as described 15 in the above cited references. Accordingly, cells and cell lines that produce monoclonal antibodies against an amino acid sequence of the invention form a further aspect of the invention, as do methods for producing antibodies against amino acid sequences of the invention, which methods may generally involve cultivating such a cell and isolating the antibodies from the culture (medium), again using techniques known per se.

20 Also, Fab-fragments against the amino acid sequences of the invention (such as F(ab)₂, Fab' and Fab fragments) may be obtained by digestion of an antibody with pepsin or another protease, reducing disulfide-linkages and treatment with papain and a reducing agent, respectively. Fab-expression libraries may, for instance, be obtained by the method of Huse et al., 1989, *Science* 245:1275-1281.

25 In another embodiment, the nucleotide sequences of the invention, the amino acid sequences of the invention, and/or a host cell or host organism that expresses such an amino acid sequence, may also be used in an assay or assay method generally (including but not limited to diagnostic assays and/or assays to determine the presence and/or absence of specific mutations and/or genetic markers, for example to determine 30 susceptibility for a condition or disease associated with such a mutation or marker), and in particular in an assay to identify and/or (further) develop compounds and/or other

factors that can modulate the (biological) activity of, and/or that can otherwise interact with, the amino acid sequences of the invention, and such uses form further aspects of the invention. As will be clear to the skilled person, in this context, the amino acid sequence of the invention will serve as a target for interaction with such a compound or factor

5 In this context, the terms “modulate”, “modulation”, “modulator” and “target” will have their usual meaning in the art, for which reference is *inter alia* made to the definitions given in WO 98/06737. Generally, a modulator is a compound or factor that can enhance, inhibit/reduce or otherwise alter, influence or affect (collectively referred to as “modulation”) a functional property of a biological activity or process (for example, 10 the biological activity of an amino acid sequence of the invention).

In this context, the amino acid sequence of the invention may serve as a target for modulation *in vitro* (e.g. as part of an assay or screen) and/or for modulation *in vivo* (e.g. 15 for modulation by a compound or factor that is known to modulate the target, which compound or factor may for example be used as an active compound for agrochemical, veterinary and/or pharmaceutical use).

For example, the amino acid sequences, host cells and/or host organisms of the invention may be used as part of an assay or screen that may be used to identify and/or develop modulators of the amino acid sequence of the invention, such as a primary screen (e.g. a screen used to identify modulators of the target from a set or library of test 20 chemicals with unknown activity with respect to the target) and/or a secondary assay (e.g. an assay used for validating hits from a primary screen and/or used in optimizing hit molecules, e.g. as part of hits-to-leads chemistry).

For instance, such an assay or screen may be configured as an *in vitro* assay or screen, which will generally involve binding of the compound or factor to be tested as a 25 potential modulator for the target (hereinbelow also referred to as “test chemical”) to the target, upon which a signal generated by said binding is measured. Suitable techniques for such *in vitro* screening will be clear to the skilled person, and are for example described in Eldefrawi et al., (1987). FASEB J., Vol.1, pages 262-271 and Rauh et al., (1990), Trends in Pharmacol. Sci., vol.11, pages 325-329. For example, such an assay or 30 screen may be configured as a binding assay or screen, in which the test chemical is used to displace a detectable ligand from the target (e.g. a radioactive or fluorescent ligand),

upon which the amount of ligand displaced from the target by the modulator is determined.

Assays aimed at identification of small molecule inhibitors of protein kinases are well known. There is a large number of technologies that can be classified in two main approaches:

5 1. Detection based on an antibody specifically recognizing the phosphorylated product (peptide) of the kinase action

These approaches rely on the specificity of antibody to differentiate between the peptide containing a phosphorylated amino acid vs. the non-phosphorylated peptide.

10 This recognition can utilize, for example, the following readouts:

- fluorescence polarization (FP), where the detection is based on the change of the overall size upon antibody binding to fluorescent labeled peptide.
- fluorescence intensity (FI), where the antibody binding causes change in the intensity of the emission from a fluorescently labeled peptide.
- time-resolve fluorescence (where both the antibody and the peptide are labeled with different fluorescent labels and the binding leads to energy transfer between these two labels. This in turn leads to shortening of the donor label fluorescence lifetime. These assays have at least two different embodiments: biochemical or cell-based, via enzyme complementation (the latter is marketed as HitHunter

15 (TM), by Discoverex) .

20 2. Detection based on other reagents specifically recognizing phosphate groups.

Recently, at least three new techniques have been developed that do not rely on the use of specific antibody.

- Trivalent metal ion containing beads. These bind specifically to phosphate groups. The detection is then performed by change in fluorescence polarization, the same as above, upon antibody binding. (IMAP(TM), Molecular Devices))
- based on chemical modification of the phosphate group and a specific label in the substrate to yield a change in fluorescence intensity. (IQ(TM), Pierce)
- based on chemical modification of the phosphate group to yield a biotin derivative. The detection is performed via FP upon avidin or strepavidin binding. (Caliper)"

Such an assay or screen may also be configured as a cell-based assay or screen, in which a host cell of the invention is contacted with/exposed to a test chemical, upon which at least one biological response by the host cell is measured.

Suitable cells or cell lines for such cell based assays include those mentioned 5 above. In one preferred, but non-limiting embodiment, the cell or cell line may be a mammalian, and in particular human, cell or cell line which is related to metabolic processes or metabolic disease and/or used as a cellular model for metabolic disease, including but not limited to liver cells or cell lines, adipocytes or muscle cells or cell lines such as HEPG2 cells, 3T3L1 adipocytes, CTC12 cells and L6 myotubes.

10 Also, such an assay or screen may also be configured as a whole animal screen, in which a host organism of the invention is contacted with/exposed to a test chemical, upon which at least one biological response (such as a phenotypical, behavioural and/or physiological change, including but not limited to paralysis or death) by the host organism is measured. Such screens may be carried out in any model organism known 15 per se, including but not limited to yeast, *Drosophila*, zebrafish or *C. elegans*.

Thus, generally, the assays and screens described above will comprise at least one step in which the test chemical is contacted with the target (and/or with a host cell or host 20 organism that expresses the target), and in particular in such a way that a signal is generated that is representative for the modulation of the target by the test chemical (including modulation of the interaction of the target protein with any other proteins. Suitable techniques for screening such protein-protein interactions, such as FRET, SPA and BRET, will be clear to the skilled person). In a further step, said signal may then be detected.

Accordingly, in one aspect, the invention relates to a method for generating a 25 signal that is representative for the interaction of an amino acid sequence of the invention with a test chemical, said method at least comprising the steps of:

- a) contacting the amino acid sequence of the invention, or a host cell or host organism containing/expressing an amino acid sequence of the invention, with said test chemical, in such a way that a signal may be generated that is representative for the 30 interaction between said test chemical and said amino acid sequence; and optionally
- b) detecting the signal that may thus be generated.

In another aspect, the invention relates to a method for identifying modulators of an amino acid sequence of the invention (e.g. from a set or library of test chemicals), said method at least comprising the steps of:

- a) contacting the amino acid sequence of the invention, or a host cell or host organism containing/expressing an amino acid sequence of the invention, with a test chemical, in such a way that a signal may be generated that is representative for the interaction between said test chemical and said amino acid sequence; and optionally
- b) detecting the signal that may thus be generated, said signal identifying a modulator of said amino acid sequence.

Compounds that may be tested using the methods of the invention are generally described below.

The assays and screens of the invention may be carried out at medium throughput to high throughput, for example in an automated fashion using suitable robotics. In particular, in this embodiment, the method of the invention may be carried out by contacting the target with the test compound in a well of a multi-well plate, such as a standard 24, 96, 384, 1536 or 3456 well plate.

Usually, in a screen or assay of the invention, for each measurement, the target or host cell/host organism will be contacted with only a single test compound. However, it is also within the scope of the invention to contact the target with two or more test compounds - either simultaneously or sequentially - for example to determine whether said combination provides a synergistic effect.

Once a test chemical has been identified as a modulator for an amino acid sequence of the invention (e.g. by means of a screen or assay as described hereinabove), it may be used per se as a modulator of the amino relevant amino acid sequence of the invention (e.g. as an active substance for pharmaceutical use), or it may optionally be further optimized for final use, e.g. to improve properties such as solubility, ADME-TOX (see for example <http://www.scherago.com/admet/> or <http://www.lifesciencesinfo.com/admet>) and other desired properties. It will be clear to the skilled person that the nucleotide sequences, amino acid sequences, host cells/host organisms and/or methods of the invention may find further use in such optimization methodology, for example as (part of) secondary assays.

The invention is not particularly limited to any specific manner or mechanism in/via which the modulator (e.g. the test chemical, compound and/or factor) modulates, or interacts with, the target (*in vivo* and/or *in vitro*). For example, the modulator may comprise an agonist, an antagonist, an inverse agonist, a partial agonist, a competitive 5 inhibitor, a non-competitive inhibitor, a cofactor, an allosteric inhibitor or other allosteric factor for the target, and/or may be a compound or factor that enhances or reduces binding of target to another biological component associated with its (biological) activity, such as another protein or polypeptide, a receptor, or a part or organelle of a cell. As such, the modulator may bind with the target (at the active site, at an allosteric site, at a 10 binding domain and/or at another site on the target, e.g. covalently or via hydrogen bonding), block the active site of the target (in a reversible, irreversible or competitive manner), block a binding domain of the target (in a reversible, irreversible or competitive manner), and/or influence or change the conformation of the target.

As such, the test chemical/modulator may for instance be:

- 15 - an analog of a known substrate of the target;
- an oligopeptide, e.g. comprising between 2 and 20, preferably between 3 and 15 amino acid residues;
- an antisense or double stranded RNA molecule;
- a protein, polypeptide;
- 20 - a cofactor or an analog of a cofactor.

Preferably, the compound is an inhibitor of the target, although the invention in its broadest sense is not limited thereto.

The test chemical/modulator may also be a reference compound or factor, which may be a compound that is known to modulate or otherwise interact with the target (e.g. a 25 known substrate or inhibitor for the target) or a compound or factor that is generally known to modulate or otherwise interact with other members from the general class to which the target belongs (e.g. a known substrate or inhibitor of said class).

Preferably, however, the compound(s) will be "small molecules", by which is generally meant herein a molecular entity with a molecular weight of less than 1500, 30 preferably less than 1000. This may for example be an organic, inorganic or organometallic molecule, which may also be in the form or a suitable salt, such as a

water-soluble salt; and may also be a complex, chelate and/or a similar molecular entity, as long as its (overall) molecular weight is within the range indicated above.

In a preferred embodiment, such a "small molecule" has been designed according, and/or meets the criteria of, at least one, preferably at least any two, more preferably at 5 least any three, and up to all of the so-called Lipinski rules for drug likeness prediction (vide Lipinski et al., Advanced Drug Delivery Reviews 23 (1997), pages 3-25). As is known in the art, small molecules which meet these criteria are particularly suited (as starting points) for the design and/or development of pharmaceuticals for human use, and may for instance be used as starting points for hits-to-leads chemistry, and/or as starting 10 points for lead development (in which the methods of the invention may also be applied).

Also, for these purposes, the design of such small molecules (as well as the design of libraries consisting of such small molecules) will preferably also take into account the presence of pharmacophore points, for example according to the methods described by I. Muegge et al., J. Med. Chem. 44, 12 (2001), pages 1-6 and the documents cited herein.

15 The term "small peptide" generally covers (oligo)peptides that contain a total of between 2 and 35, such as for example between 3 and 25, amino acids (e.g. in one or more connected chains, and preferably a single chain). It will be clear that some of these small peptides will also be included in the term small molecule as used herein, depending on their molecular weight.

20 In one preferred, but non-limiting embodiment, the invention is used to screen a set or library of (related or otherwise unrelated) small molecules, for example a standard "robustness set", a primary screening library (e.g. of otherwise unrelated compounds), a combinatorial library, a series of closely related chemical analogos. Such sets or libraries will be clear to the skilled person, and may for instance include, but are not limited to, 25 such commercially available chemical libraries such as the various libraries available from Tocris Cookson, Bristol, UK.

30 In yet a further aspect, the invention relates to a method for identifying a compound that can be used in (the preparation of a pharmaceutical composition for) the prevention and/or treatment of metabolic diseases (e.g. from a set or library of test chemicals), said method at least comprising the steps of:

- a) contacting an amino acid sequence of the invention, and/or a host cell or host organism containing/expressing an amino acid sequence of the invention, with a test chemical, in such a way that a signal may be generated that is representative for the interaction between said test chemical and said amino acid sequence of the invention;
- 5 and optionally
- b) detecting the signal that may thus be generated, said signal identifying a modulator of said amino acid sequence.

The modulators thus identified can be used in (the preparation of a pharmaceutical composition for) the prevention and/or treatment of metabolic diseases, and/or can be
10 used to develop other compounds that can be used in (the preparation of a pharmaceutical composition for) the prevention and/or treatment of metabolic diseases, i.e. as already outlined above.

The invention also relates to the use of an amino acid sequence and/or a host cell/host organism of the invention in such a method.

15 Also, as already mentioned above, the use of the human nucleotide sequences of SEQ ID NOS: 3, 5, 7 and/or 9 and/or of sequences derived therefrom (such as mutants, parts, fragments and/or fusions thereof as described hereinabove), of the human amino acid sequences of SEQ ID NOS: 4, 6, 8 and/or 10 and/or of sequences derived therefrom (such as analogs, parts, fragments, and/or fusions thereof as described hereinabove), and
20 of host cells/host organisms containing/expressing these, are usually preferred, in particular when the invention is used to develop compounds for pharmaceutical use.

25 As already mentioned above, the compounds and/or factors that have been identified and/or developed as modulators of the amino acid sequences of the invention (and/or precursors for such compounds) may be useful as active substances in the pharmaceutical field, for example in the preparation of pharmaceutical compositions, and both such modulators as well as (pharmaceutical) compositions containing them comprise further aspects of the invention.

30 In particular, the compounds and composition of the invention may be used in (the preparation of pharmaceutical compositions for) the prevention (e.g. prophylaxis) and/or treatment of metabolic diseases (which for the purposes herein in its broadest

sense also includes preventing, treating and/or alleviating the symptoms and/or complications of such metabolic diseases).

In particular, such compounds and composition may be used in (the preparation of pharmaceutical compositions for) the prevention (e.g. prophylaxis) and/or treatment of metabolic diseases (which for the purposes herein in its broadest sense also includes preventing, treating and/or alleviating the symptoms and/or complications of such metabolic diseases).

In particular, the compounds and compositions of the invention may be used for preventing and/or treating:

- 10 - hyperglycemic conditions and/or other conditions and/or diseases that are (primarily) associated with (the response or sensitivity to) insulin, including but not limited to all forms of diabetes and disorders resulting from insulin resistance, such as Type I and Type II diabetes, as well as severe insulin resistance, hyperinsulinemia, and hyperlipidemia, e.g., obese subjects, and insulin-resistant diabetes, such as
- 15 Mendenhall's Syndrome, Werner Syndrome, leprechaunism, lipoatrophic diabetes, and other lipoatrophies;
- conditions caused or usually associated with hyperglycemic conditions and/or obesity, such as hypertension, osteoporosis and/or lipodystrophy.
- so-called "metabolic syndrome" (also known as "Syndrome X") which is a condition
- 20 where several of the following conditions coexist: hypertension; insulin resistance; diabetes; dyslipidemia; and/or obesity.

In particular, the compounds and compositions of the invention may be used for preventing and/or treating diabetes, especially Type I and Type II diabetes. "Diabetes" itself refers to a progressive disease of carbohydrate metabolism involving inadequate production or utilization of insulin and is characterized by hyperglycemia and glycosuria.

Also, as mentioned above, the amino acid sequences of the invention and in particular the nucleotide sequences of the invention, and more in particular the human amino acid sequences and nucleotide sequences of the invention may be used for diagnostic purposes, for example as part of diagnostic assays and/or as part of kits for performing such assays (in which such a kit will comprise at least a nucleotide sequence of the invention) may be suitably packaged (e.g. in a suitable container) and may

optionally further comprise one or more elements for such kits known per se, such as suitable reagents, buffers or other solvents, and instructions for use).

In particular, the amino acid sequences and nucleotide sequences of the invention, as well as assays and kits using such sequences, may be used for diagnostic purposes

5 relating to one or more of the metabolic diseases specified, for example as assays to determine the presence and/or absence in an individual of specific mutations and/or genetic markers that relate to one or more of the metabolic diseases specified above, to determine the susceptibility and/or any predisposition to any of said metabolic diseases in an individual, to determine if any genetically determined factors contribute or even cause

10 (in full or in part) said metabolic disease in an individual, determine and/or to confirm the kind of metabolic disease from which an individual suffers, and/or to predict the further progress of such a metabolic disease in an individual. It will also be clear that any results obtained using such a diagnostic method or assay may also provide guidance to the clinician as to how such a metabolic disease should be treated in an individual, e.g. which

15 medication should be prescribed and/or the dosage regimen to be used.

It should also be noted that, for the treatment of metabolic disease in humans, the compound used will usually and preferably be an inhibitor of an amino acid sequence of the invention, although the invention in its broadest sense is not limited thereto.

In one specific, but non-limiting, embodiment of the invention, a compound is

20 considered an inhibitor of one of the amino acid sequences of the invention if, in a relevant assay such as the kinase activity assays referred to above (or a suitable modification thereof, for example using partially or fully purified protein), said compound reduces the activity of said amino acid sequence, i.e. by at least 1%, preferably at least 10%, such as by 20% or more, compared to the activity without the presence of

25 said compound.

In an even more specific, but non-limiting, embodiment of the invention, a compound is considered an inhibitor of one of the amino acid sequences of the invention if, in a relevant assay, such as a binding assay, said compound has an IC₅₀ value of less than 1000 μ M, preferably at less than 500 μ M, more preferably less than 250 μ M, even more

30 preferably less than 100 μ M, for example 50 μ M or less, such as about 10 μ M or less.

Again, preferably, in the invention compounds are used that are modulators, and in particular inhibitors, of the human amino acid sequences of SEQ ID NO: 4, 6, 8 and/or 10, and/or of amino acid sequences derived therefrom, such as analogs, mutants, parts, fragments and/or fusions as described above.

5 For pharmaceutical use, the compounds of the invention may be used as a free acid or base, and/or in the form of a pharmaceutically acceptable acid-addition and/or base-addition salt (e.g. obtained with non-toxic organic or inorganic acid or base), in the form of a hydrate, solvate and/or complex, and/or in the form or a pre-drug, such as an ester. Such salts, hydrates, solvates, etc. and the preparation thereof will be clear to the
10 skilled person; reference is for instance made to the salts, hydrates, solvates, etc. described in US-A-6,372,778, US-A-6,369,086 and US-6,369,067

Generally, for pharmaceutical use, the compounds of the inventions may be formulated as a pharmaceutical preparation comprising at least one compound of the invention and at least one pharmaceutically acceptable carrier, diluent or excipient and/or
15 adjuvant, and optionally one or more further pharmaceutically active compounds. By means of non-limiting examples, such a formulation may be in a form suitable for oral administration, for parenteral administration (such as by intravenous, intramuscular or subcutaneous injection or intravenous infusion), for topical administration, for administration by inhalation, by a skin patch, by an implant, by a suppository, etc. Such
20 suitable administration forms - which may be solid, semi-solid or liquid, depending on the manner of administration - as well as methods and carriers for use in the preparation thereof, will be clear to the skilled person; reference is again made to for instance US-A-6,372,778, US-A-3,696, 086 and US-6,369,067.

25 The pharmaceutical preparations of the invention are preferably in a unit dosage form, and may be suitably packaged, for example in a box, blister, vial, bottle, sachet, ampoule or in any other suitable holder or container (which may be properly labeled); optionally with one or more leaflets containing product information and/or instructions for use. Generally, such unit dosages will contain between 1 and 500 mg of the at least one compound of the invention, e.g. about 10, 25, 50, 100, 200, 500 or 1000 mg per unit
30 dosage.

For pharmaceutical use, at least one compound of the invention will generally be administered in an amount of between 0.01 to 150 mg/kg body weight per day of the patient, divided over one or more daily doses. The amount(s) to be administered and the further treatment regimen may be determined by the treating clinician, depending on 5 factors such as the age, gender and general condition of the patient and the nature and severity of the disease/symptoms to be treated.

Thus, in a further aspect, the invention relates to a composition, and in particular a composition for pharmaceutical use, that contains at least one compound of the invention (i.e. a compound that has been identified, discovered and/or developed using a nematode 10 or method as described herein) and at least one suitable carrier (i.e. a carrier suitable for pharmaceutical use). The invention also relates to the use of a compound of the invention in the preparation of such a composition.

Preferably, the compounds and compositions of the invention are administered orally and/or in a form intended and/or suitable for oral administration.

15 It is also envisaged that the above compounds and compositions may be of value in the veterinary field, which for the purposes herein not only includes the prevention and/or treatment of diseases in animals, but also - for economically important animals such as cattle, pigs, sheep, chicken, fish, etc. - enhancing the growth and/or weight of the animal and/or the amount and/or the quality of the meat or other products obtained from 20 the animal. Thus, in a further aspect, the invention relates to a composition for veterinary use that contains at least one compound of the invention (i.e. a compound that has been identified, discovered and/or developed using a nematode or method as described herein) and at least one suitable carrier (i.e. a carrier suitable for veterinary use). The invention also relates to the use of a compound of the invention in the preparation of such a 25 composition.

In the agrochemical field, the invention may be used to identify compounds suitable for use in pesticides, insecticides, nematicides and/or other biocides or plant protection agents. For example, the compounds of the invention may be used to control the species listed in US-A-6,372,774. For this purpose, the compounds of the invention 30 (or a suitable salt, hydrate or ester thereof) may be suitably formulated with one or more agrochemically acceptable carriers, to provide a formulation suitable for agrochemical

use, as will be clear to the skilled person (reference is, for example, made to the formulations and uses described in US-A-6,372,774).

Thus, in a further aspect, the invention relates to a composition for agrochemical use that contains at least one compound of the invention (i.e. a compound that has been 5 identified, discovered and/or developed using a nematode or method as described herein) and at least one suitable carrier (i.e. a carrier suitable for agrochemical use). The invention also relates to the use of a compound of the invention in the preparation of such a composition.

The invention will now be further illustrated by means of the following non-10 limiting Experimental Part.

In the Figures:

- Figure 1 schematically shows vector pGN49A (see also also WO 00/01846 and British patent application 0012233, both by Applicant);
- Figures 2A and 2B are photographs (enhanced using the Scion Image (Scion Corp) 15 software package) showing reduced fat-absorption phenotype in *C. elegans* upon Nile Red Staining: Figure 2A = reduced fat storage (T17E9.1a- invention); Figure 2B = reference (vector gGN29 without T17E9.1a).

Experimental part:

20 In the Experimental Part below, unless indicated otherwise, all steps for handling and cultivating *C. elegans* were performed using standard techniques and procedures, for which reference is made to the standard *C. elegans* handbooks, such as W.B. Wood et al., "The nematode *Caenorhabditis elegans*", Cold Spring Harbor Laboratory Press (1988); D.L. Riddle et al., "C. ELEGANS II", Cold Spring Harbor Laboratory Press (1997); 25 "Caenorhabditis elegans, Modern Biological analysis of an organism": ed. by H. Epstein and D. Shakes, Methods in Cell Biology, Vol 48, 1995; and "C. elegans, a practical approach", ed. by I.A. Hope, Oxford University Press Inc. New York, USA, 1999.

30 Downregulation of the gene(s) of interest in *C. elegans* was achieved by RNAi feeding techniques using an *E.coli* strain capable of expressing a dsRNA corresponding to the gene(s) of interest, as generally described in - *inter alia* - the International application WO 00/01846 by applicant and the handbooks referred to above.

Also, unless indicated otherwise, all cloning and other molecular biology steps were performed using standard techniques and protocols, i.e. as provided by the manufacturers of the reagents/kits used and/or as described in the standard handbooks, such as Sambrook et al, "Molecular Cloning: A Laboratory Manual" (2nd.ed.), Vols. 1-3, 5 Cold Spring Harbor Laboratory Press (1989) and F. Ausubel et al, eds., "Current protocols in molecular biology", Green Publishing and Wiley Interscience, New York (1987).

Fat accumulation in *C. elegans daf-2 (e1370)* was determined visually under a microscope upon staining with Nile-red, using an adaptation of the general methodology 10 described by Ogg et al., *Nature*, Vol. 389, 994 (1997). For the general methodology, reference is also made to Thaden et al., 1999 International Worm Meeting abstract 837; Ashrafi and Ruvkun, 2000 East Coast Worm Meeting abstract 67; Ashrafi, Chang and Ruvkun, 2001 International Worm Meeting abstract 325; and Rottiers and Antebi, 2001 International Worm Meeting abstract 620 (all abstracts available from Worm Literature 15 Index at <http://elegans.swmed.edu/wli/>).

Example 1: Preparation *E.coli* RNA feeding strain for expression of T17E9.1a double stranded RNA.

A vector for expression of dsRNA for downregulation of *C. elegans* gene 20 T17E9.1a (*kin-18*) was prepared as follows.

The DNA fragment of SEQ ID NO:11, which corresponds to 1576 nucleotides (26 % compared to the genomic sequence) of the *C. elegans* T17E9.1a gene, was obtained by PCR from genomic *C. elegans* DNA, using the following primers:

25 - forward primer : CGAAAACCAGCAGAAGAGCG [SEQ ID NO: 12]
- reverse primer : TCGAGGCATGTTCAGACTTCG [SEQ ID NO: 13]

This fragment was inserted in the *SrfI*-site of expression vector pGN49a pGN49A (Figure 1, see also WO 00/01846 and British patent application 0012233, both by 30 Applicant). This vector contains two T7 promoters flanking the *SrfI*-site, allowing

transcription of a nucleotide sequence inserted into said *SrfI*-site into double stranded RNA, upon binding of a T7 polymerase to said promoter (vide WO 00/01846).

The resulting vector, designated pGN49A-T17E9.1a, was transformed overnight into *E. coli* strain AB 309-105 (see EP-A-1 093 526 by applicant, page 17.).

5 To normalize the culture, 250 μ l of the overnight culture (1 ml) was transferred to a 96 well plate and the OD at 600 nm was measured (Fluostar Galaxy plate reader BMG), the remaining 750 μ l centrifuged down. Next the pellet was re-suspended in S-complete fed (S-complete supplemented with 0.1mg/ml ampicilin and 1 mM IPTG) and volume adjusted to obtain OD₆₀₀ value of 1

10

Example 2: Generation of fat storage phenotype in *C. elegans*.

In this example, *C. elegans* strain CB1370 containing the temperature sensitive daf-2 allele e-1370 is used (Ogg et al., supra). CB 1370 is publicly available from, for example, the Caenorhabditis Genetics Center (CGC), Minnesota, USA).

15

To generate the fat-storage phenotype, L1 worms of strain CB 1370 were cultivated at a temperature of 15 °C in S-Complete fed-medium in the wells of a 96 well plate (1 L1 nematodes per well) under essentially synchronized conditions, until the nematodes reached the L2 stage.

20

Then, the temperature was increased to 20°C, and the worms were further cultivated at said temperature until their F1 offspring reached the L4 stage (about 144 hrs). Due to the presence of the daf-2 allele e-1370, this raise in temperature from 15°C to 20°C causes the nematodes to accumulate fat, mainly in their intestinal and hypodermal tissue (vide Ogg et al. and Figures 2A and 2B).

25

The accumulation of fat (in the form of droplets) was made visible by means of Nile Red staining: L4 animals were washed several times with M9 (supplemented with 0.1% PEG) to remove the remaining *E. coli*, and fixed with MeOH (fc. 33%). After fixation the nematodes were stained with nile red (fc 0.375 mM in 37.5% MeOH) for 4 hours. MeOH and excess dye was removed through several washes with M9 (supplemented with 0.1% PEG). The staining pattern was visualized under UV using a 30 500 nm long pass filter.

For testing the influence of the gene T17E9.1a on fat storage, during the steps described above, the worms were grown on 30 μ l of the normalized *E. coli* pGN49A-T17E9.1a strain of Example 1 ($OD_{600} = 1$) as a food source. As a reference, the daf-2 (e1370) nematodes were grown in a similar manner, but with *E. coli* strain AB 309-105 5 containing vector pGN29A without the T17E9.1a insert as a food source, used in the same amount. All samples were carried out in quadruplicate.

The results were as follows: worms fed on *E. coli* pGN49A-T17E9.1a strain, which downregulates the expression of T17E9.1a through RNA interference, showed a strong reduction of the accumulation of fat, compared to the reference (vide Figures 2A 10 and 2B).

These results show that T17E9.1a is involved in the regulation of the (daf-2 dependent) *daf*-pathway and the accumulation of fat in the nematode. It is known in the art that both are models for insulin resistance and fat handling in mammals, such as humans.

15 Bio-informatic analysis of the sequence of T17E9.1a (kin-18) showed that the amino acid sequence of the kin-18 protein (SEQ ID NO:2) has the following degree of sequence identity (in percentages at the amino acid level) with the human amino acid sequences of SEQ ID NOS: 4, 6, 8 and 10, respectively:

20 - SEQ ID NO:4 ("JIK"): overall sequence: 36%; kinase domain: 61%
- SEQ ID NO:6 ("PSK"): overall sequence: 24%; kinase domain: 59%
- SEQ ID NO:8 ("TAOI"): overall sequence: 32%; kinase domain: 59%
- SEQ ID NO:10 ("Q9P2I6"): overall sequence: 33%; kinase domain: 61%

Example 3: Screening of SEQ ID NO. 4 ("JIK").

A protein with the amino acid sequence of SEQ ID NO: 4, isolated and purified in a manner known per se, were used to screen a reference set of 96 commercially available kinase inhibitors (obtained from several commercial sources such as Acros Organics,

5 Belgium/USA; Alexis Biochemicals, UK; Biomol, USA; Calbiochem, USA; ICN Biomedicals; Sigma; and Tocris, UK), using the commercially available Kinase Glo® Luminescent Kinase Assay (Promega, USA), essentially according to the instructions of the manufacturer.

In this assay, 6 out of 96 compounds showed a percentage which differed more
10 than 3 Standard Deviation (SD) from 0 (zero) inhibition. These 6 compounds were designated as hits, corresponding with a hit-rate of about 6 %.

Example 4: Screening of SEQ ID NO. 8 ("TAOI").

A protein with the amino acid sequence of SEQ ID NO: 8, isolated and purified in a manner known per se, were used to screen a reference set of 96 commercially available kinase inhibitors (obtained from several commercial sources such as Acros Organics, Belgium/USA; Alexis Biochemicals, UK; Biomol, USA; Calbiochem, USA; ICN Biomedicals; Sigma; and Tocris, UK), using the commercially available Kinase Glo® Luminescent Kinase Assay (Promega, USA), essentially according to the instructions of the manufacturer.

In this assay, 6 out of 96 compounds showed a percentage which differed more than 3 Standard Deviation (SD) from 0 (zero) inhibition. These 6 compounds were designated as hits, corresponding with a hit-rate of about 6 %.

CLAIMS:

5 1. Nucleic acid, in essentially isolated form, said nucleic acid encoding one of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10, or an analog, variant, allele, ortholog, part and/or fragment of one of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10.

10 2. Nucleic acid, in essentially isolated form, said nucleic acid comprising one of the nucleotide sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 and/or SEQ ID NO: 9, or a mutant, variant, allele, analog, ortholog, part and/or fragment thereof.

15 3. Genetic construct, comprising the nucleic acid of claim 1 and/or of claim 2, and optionally one or more further elements of genetic constructs known per se.

20 4. Host cell or host organism, transformed with and/or containing a nucleic acid according to claim 1 and/or claim 2 and/or a genetic construct according to claim 3.

25 5. Host cell or host organism, (capable of) expressing and/or producing one of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10 or an analog, variant, allele, analog, ortholog, part and/or fragment thereof.

30 6. Method for producing one of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10, or an analog, variant, allele, ortholog, part and/or fragment thereof, said method comprising at least the steps of:

35 a) expressing of a nucleic acid according to claim 1 and/or 2 in a suitable host cell or host organism; and optionally

b) isolating the amino acid sequence thus expressed.

7. Method for producing one of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10, or an analog, 5 variant, allele, ortholog, part and/or fragment thereof, said method comprising at least the steps of:

a) maintaining and/or cultivating a host cell or host organism according to claim 4 or 5 under conditions such that said host cell or host organism expresses or produces one 10 of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10, or an analog, variant, allele, ortholog, part and/or fragment thereof; and optionally

b) isolating the amino acid sequence thus expressed/produced.

8. Protein or polypeptide, in essentially isolated form, said protein or polypeptide 15 comprising one of the amino acid sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and/or SEQ ID NO: 10, or an analog, variant, allele, ortholog, part and/or fragment thereof.

9. Method for generating a signal that is representative for the interaction of a 20 protein or polypeptide of claim 8 with a test chemical, said method at least comprising the steps of:

a) contacting a protein or polypeptide of claim 8, or a host cell or host organism of claim 4 or 5, with said test chemical, in such a way that a signal may be generated that is representative for the interaction between said test chemical and said protein or 25 polypeptide; and optionally

b) detecting the signal that may thus be generated.

10. Method for identifying a modulator of a protein or polypeptide of claim 8, for example from a set or library of test chemicals, said method at least comprising the steps 30 of:

- a) contacting a protein or polypeptide of claim 8, or a host cell or host organism of claim 4 or 5, with said test chemical, in such a way that a signal may be generated that is representative for the interaction between said test chemical and said protein or polypeptide; and optionally
- 5 b) detecting the signal that may thus be generated, said signal identifying a modulator of said amino acid sequence.

10 11. Modulator of a protein or polypeptide according to claim 8, identified and/or developed using a nucleic acid according to claim 1 and/or 2, a host cell according to claim 4 and/or 5, a protein or polypeptide according to claim 8, and/or a method according to claim 9 and/or 10.

15 12. Modulator according to claim 11, being an inhibitor of a protein or polypeptide of claim 8.

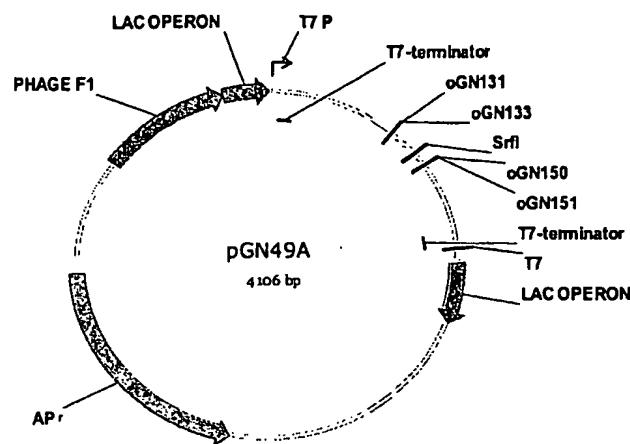
13. Pharmaceutical composition, comprising at least one modulator according to claim 11 and/or 12 and at least one pharmaceutically acceptable carrier.

20 14. Pharmaceutical composition according to claim 13, being a composition suitable and/or intended for oral administration.

15. Use of a modulator according to claim 11 and/or 12 in the preparation of a pharmaceutical composition.

25 16. Use of a modulator according to claim 11 and/or 12 in the preparation of a pharmaceutical composition for the prevention and/or treatment of metabolic diseases.

17. Antibody against a protein or polypeptide according to claim 8.

Figure 1

10/540634

WO 2004/056982

PCT/EP2003/014674

2/2

Figure 2A

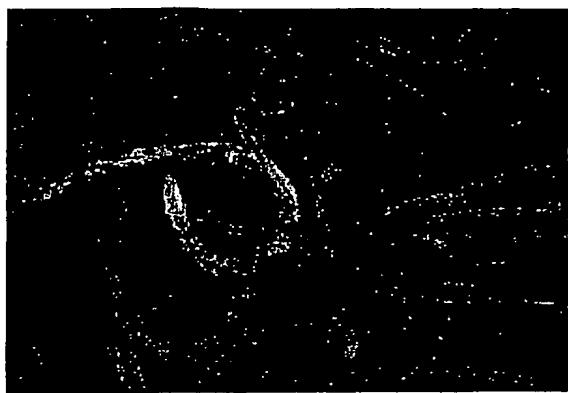


Figure 2B



SEQUENCE LISTING

5 <110> Devgen N.V.
5 Devgen N.V.

10 <120> Kinase sequences useful for developing compounds for the
prevention and/or treatment of metabolic diseases and nucleotide
sequences encoding such kinase sequences.

15 <130> P 02-011

15 <160> 13

15 <170> PatentIn version 3.1

20 <210> 1

20 <211> 2949

20 <212> DNA

25 <213> *Caenorhabditis elegans*

25	<400> 1	
	atggcgcctg ccgtcttaca aaaacccggt gttatcaagg atccatcgat tgctgcattg	60
30	ttcagtaata aggatccaga gcagagatat caagatttaa gagaaattgg acatggatct	120
	tttggagctg tctatttgc atatgacaaa aaaaatgagc agactgttgc gattaaaaag	180
35	atgaatttta gtggaaaaca ggctgtcgaa aaatggaatg atattcttaa agaagtgtct	240
	tttctgaata cagttgttca tccacatatt gtcgactaca aggcttgtt tcttaaggac	300
	actacatgtt ggcttgtat ggagtactgt attggctctg cagccgatata agtggatgtc	360
40	ttgcgaaaag gaatgcgaga agtcgaaatc gctgcgattt gctctcaaac tttggatgtc	420
	cttcgatatac ttcactctct gaagcgaata catcgagata ttaaagctgg aaatattctg	480
45	ctatctgatc atgctattgt taaactagct gatttcgat ccgcattccct ggtagatccg	540
	gctcaaactt tcatcggaac gccgttttc atggccccag aggttaattct ggcaatggat	600
	gagggtcact acacggatcg tgcagatatt tggtcattgg gtatcacgtg tatagagctg	660
50	gccgaacgtc gtccaccatt gttcagtatg aatgcaatgt ctgcctcta ccatattgt	720
	caaaatgatc ctccaactct ttctccaatt gacactagcg aacaaccgga atggtcgctg	780
55	gaattcgttc aatttataga caaatgtctt cgaaaaccag cagaagagcg aatgtcagct	840
	gaagaatgct ttcgacatcc attcattcaa cggctcgcc catcagacac aattcaggaa	900

	ctcattcaga gaacgaaaaa tatggtatta gagttggata atttcaata caaaaagatg	960
	agaaaaactca tgtatggaa tgaaacagaa gggaaaagaag gaagtgaagg aatggagca	1020
5	tctgatgatt tagatttca tggaaatgaa gctaattcaa ttggaaagagc aggagattct	1080
	gcgtcatctc gaagtgcctc tcttacttct ttccgatcaa tgcagagtag tggaggagct	1140
10	ggtcttttag tgtccaccaa tacgacgggt gctatggata atgtgcatttgc atcctctgga	1200
	tacggtaatg gaagttagttc gacgacgagc tccgcacgccc gccgtcctcc aattccttcg	1260
	caaattgtctt cttctacatc aacgtctgggt gttggaaacta tgccgagtca tggatcagg	1320
15	ggagcatcgat ttacggcgat cgccgtcaat ccaacacccgt ctccattcaga acctatccca	1380
	acatcacaac caacatcgaa atcagaatca tcttctatac tcgaaactgc acacgatgat	1440
20	cctttggaca cgtcgataacg tgctccagtg aaagacttgc atatgccgca tcgaggcgtc	1500
	aaggaacgaa tagccacgtt gcaaaatcac aaattcgcga cgcttcgttc ccagagaata	1560
	atcaatcagg aacaagaaga atatacgaaa gagaacaata tgtatgagca aatgagcaag	1620
25	tacaaggcatc tacgacaagc acatcacaaa gagctccaac aatttgaaga acgtatgtca	1680
	ttagatagag agcaactgcg tgtgaaaatg gatcgagaac tcgaaacaattt gacaacgaca	1740
30	tactcgaaag aaaagatgag agtgagggtgt tcacagaata atgaactaga caaacggaaa	1800
	aaagatatcg aagatggggaa gaaaaagatg aaaaagacga aaaatagtca aatcagcag	1860
	cagatgaaac tgtatttcagc gcaacaattt aagaataca agtataacaa ggaggcacag	1920
35	aaaacacgat tacgaagtct gaacatgcct cgaagtactt atgagaacgc aatgaaagaa	1980
	gtgaaaagccg atctgaatcg agtggaaatg gcaacggggaaa atgattttga cgagaagctt	2040
40	cgtgcagaac ttgaagatga aattttaagg tatcgccaggc aacaactcag taatcttcatt	2100
	caatttggaaag aacaatttggaa tgatgaagac gttaaacgtgc aagaacgcca aatggacacg	2160
	cgtcacggat tactgtcaaa gcagcatgaa atgacgcgcg atttggaaat acagcatctc	2220
45	aacgagcttc acgcgtatgaa aaaacgacat ttggagacac aacacgaggc ggaatcggca	2280
	agtcaaaaatg agtacacaca gaggcaacag gatgaatttga gaaaaaaagca tgcgatgcag	2340
50	tcaagacaac agccaagaga tttaaagatc caagaagcac aaattcgaaa acaataccga	2400
	caagttgtga agactcagac tcgccaattt aagctctacc ttacacaaat ggtgcaagta	2460
	gttccaaaag atgaacaaaa agagctcagc tctcgactaa aacaggatca aatgcaaaaa	2520
55	gtcgacttc ttgccttcaca atacgaaatg caaatcaaaa aaatggttca ggataagaca	2580
	gtgaagctcg agtcgtggca agaagatgaa caacgggttc ttagtgagaa gttggagaaa	2640

5 gaatttggaaag aattgattgc ttatcagaag aagacgagag ccacattaga agagcagatt 2700
 aaaaaggaac gtacggcact cgaagaacga attggcacac gacgtgcaat gcttgaacag 2760
 aagattattg aagaacgcga acaaatggga gaaatgcgtc gactaaagaa ggagcaaatc 2820
 cgtgatcgac acagtcaaga acgccccatcgt ctcgagaatc atttcgtacg gacgggctcg 2880
 10 acgagcagaa gttctgggtgg gatcgctcct ggtgttggga attcaaggcag tattcagatg 2940
 gctatgttag 2949

15 <210> 2
 <211> 982
 <212> PRT
 20 <213> *Caenorhabditis elegans*
 <400> 2

25 Met Ala Pro Ala Val Leu Gln Lys Pro Gly Val Ile Lys Asp Pro Ser
 1 5 10 15
 Ile Ala Ala Leu Phe Ser Asn Lys Asp Pro Glu Gln Arg Tyr Gln Asp
 20 25 30
 30 Leu Arg Glu Ile Gly His Gly Ser Phe Gly Ala Val Tyr Phe Ala Tyr
 35 40 45
 35 Asp Lys Lys Asn Glu Gln Thr Val Ala Ile Lys Lys Met Asn Phe Ser
 50 55 60
 Gly Lys Gln Ala Val Glu Lys Trp Asn Asp Ile Leu Lys Glu Val Ser
 65 70 75 80
 40 Phe Leu Asn Thr Val Val His Pro His Ile Val Asp Tyr Lys Ala Cys
 85 90 95
 Phe Leu Lys Asp Thr Thr Cys Trp Leu Val Met Glu Tyr Cys Ile Gly
 100 105 110
 45 Ser Ala Ala Asp Ile Val Asp Val Leu Arg Lys Gly Met Arg Glu Val
 115 120 125
 50 Glu Ile Ala Ala Ile Cys Ser Gln Thr Leu Asp Ala Leu Arg Tyr Leu
 130 135 140
 His Ser Leu Lys Arg Ile His Arg Asp Ile Lys Ala Gly Asn Ile Leu
 145 150 155 160
 55 Leu Ser Asp His Ala Ile Val Lys Leu Ala Asp Phe Gly Ser Ala Ser
 165 170 175

Leu Val Asp Pro Ala Gln Thr Phe Ile Gly Thr Pro Phe Phe Met Ala
 180 185 190
 5 Pro Glu Val Ile Leu Ala Met Asp Glu Gly His Tyr Thr Asp Arg Ala
 195 200 205
 Asp Ile Trp Ser Leu Gly Ile Thr Cys Ile Glu Leu Ala Glu Arg Arg
 210 215 220
 10 Pro Pro Leu Phe Ser Met Asn Ala Met Ser Ala Leu Tyr His Ile Ala
 225 230 235 240
 Gln Asn Asp Pro Pro Thr Leu Ser Pro Ile Asp Thr Ser Glu Gln Pro
 245 250 255
 15 Glu Trp Ser Leu Glu Phe Val Gln Phe Ile Asp Lys Cys Leu Arg Lys
 260 265 270
 20 Pro Ala Glu Glu Arg Met Ser Ala Glu Glu Cys Phe Arg His Pro Phe
 275 280 285
 Ile Gln Arg Ser Arg Pro Ser Asp Thr Ile Gln Glu Leu Ile Gln Arg
 290 295 300
 25 Thr Lys Asn Met Val Leu Glu Leu Asp Asn Phe Gln Tyr Lys Lys Met
 305 310 315 320
 Arg Lys Leu Met Tyr Leu Asp Glu Thr Glu Gly Lys Glu Gly Ser Glu
 325 330 335
 30 Gly Asn Gly Ala Ser Asp Asp Leu Asp Phe His Gly Asn Glu Ala Asn
 340 345 350
 35 Ser Ile Gly Arg Ala Gly Asp Ser Ala Ser Ser Arg Ser Ala Ser Leu
 355 360 365
 Thr Ser Phe Arg Ser Met Gln Ser Ser Gly Gly Ala Gly Leu Leu Val
 370 375 380
 40 Ser Thr Asn Thr Thr Gly Ala Met Asp Asn Val His Gly Ser Ser Gly
 385 390 395 400
 Tyr Gly Asn Gly Ser Ser Ser Thr Thr Ser Ser Ala Arg Arg Arg Pro
 405 410 415
 45 Pro Ile Pro Ser Gln Met Leu Ser Ser Thr Ser Thr Ser Gly Val Gly
 420 425 430
 Thr Met Pro Ser His Gly Ser Val Gly Ala Ser Ile Thr Ala Ile Ala
 50 435 440 445
 Val Asn Pro Thr Pro Ser Pro Ser Glu Pro Ile Pro Thr Ser Gln Pro
 450 455 460
 55 Thr Ser Lys Ser Glu Ser Ser Ser Ile Leu Glu Thr Ala His Asp Asp
 465 470 475 480

Pro Leu Asp Thr Ser Ile Arg Ala Pro Val Lys Asp Leu His Met Pro
 485 490 495
 His Arg Ala Val Lys Glu Arg Ile Ala Thr Leu Gln Asn His Lys Phe
 5 500 505 510
 Ala Thr Leu Arg Ser Gln Arg Ile Ile Asn Gln Glu Gln Glu Glu Tyr
 515 520 525
 10 Thr Lys Glu Asn Asn Met Tyr Glu Gln Met Ser Lys Tyr Lys His Leu
 530 535 540
 Arg Gln Ala His His Lys Glu Leu Gln Gln Phe Glu Glu Arg Cys Ala
 545 550 555 560
 15 Leu Asp Arg Glu Gln Leu Arg Val Lys Met Asp Arg Glu Leu Glu Gln
 565 570 575
 Leu Thr Thr Tyr Ser Lys Glu Lys Met Arg Val Arg Cys Ser Gln
 20 580 585 590
 Asn Asn Glu Leu Asp Lys Arg Lys Lys Asp Ile Glu Asp Gly Glu Lys
 595 600 605
 25 Lys Met Lys Lys Thr Lys Asn Ser Gln Asn Gln Gln Met Lys Leu
 610 615 620
 Tyr Ser Ala Gln Gln Leu Lys Glu Tyr Lys Tyr Asn Lys Glu Ala Gln
 625 630 635 640
 30 Lys Thr Arg Leu Arg Ser Leu Asn Met Pro Arg Ser Thr Tyr Glu Asn
 645 650 655
 Ala Met Lys Glu Val Lys Ala Asp Leu Asn Arg Val Lys Asp Ala Arg
 35 660 665 670
 Glu Asn Asp Phe Asp Glu Lys Leu Arg Ala Glu Leu Glu Asp Glu Ile
 675 680 685
 40 Val Arg Tyr Arg Arg Gln Gln Leu Ser Asn Leu His Gln Leu Glu Glu
 690 695 700
 Gln Leu Asp Asp Glu Asp Val Asn Val Gln Glu Arg Gln Met Asp Thr
 705 710 715 720
 45 Arg His Gly Leu Leu Ser Lys Gln His Glu Met Thr Arg Asp Leu Glu
 725 730 735
 Ile Gln His Leu Asn Glu Leu His Ala Met Lys Lys Arg His Leu Glu
 50 740 745 750
 Thr Gln His Glu Ala Glu Ser Ala Ser Gln Asn Glu Tyr Thr Gln Arg
 755 760 765
 55 Gln Gln Asp Glu Leu Arg Lys Lys His Ala Met Gln Ser Arg Gln Gln
 770 775 780

Pro Arg Asp Leu Lys Ile Gln Glu Ala Gln Ile Arg Lys Gln Tyr Arg
785 790 795 800

5 Gln Val Val Lys Thr Gln Thr Arg Gln Phe Lys Leu Tyr Leu Thr Gln
805 810 815

Met Val Gln Val Val Pro Lys Asp Glu Gln Lys Glu Leu Thr Ser Arg
820 825 830

10 Leu Lys Gln Asp Gln Met Gln Lys Val Ala Leu Leu Ala Ser Gln Tyr
835 840 845

Glu Ser Gln Ile Lys Lys Met Val Gln Asp Lys Thr Val Lys Leu Glu
850 855 860

15 Ser Trp Gln Glu Asp Glu Gln Arg Val Leu Ser Glu Lys Leu Glu Lys
865 870 875 880

Glu Leu Glu Glu Leu Ile Ala Tyr Gln Lys Lys Thr Arg Ala Thr Leu
20 885 890 895

Glu Glu Gln Ile Lys Lys Glu Arg Thr Ala Leu Glu Glu Arg Ile Gly
900 905 910

25 Thr Arg Arg Ala Met Leu Glu Gln Lys Ile Ile Glu Glu Arg Glu Gln
915 920 925

Met Gly Glu Met Arg Arg Leu Lys Lys Glu Gln Ile Arg Asp Arg His
930 935 940

30 Ser Gln Glu Arg His Arg Leu Glu Asn His Phe Val Arg Thr Gly Ser
945 950 955 960

Thr Ser Arg Ser Ser Gly Gly Ile Ala Pro Gly Val Gly Asn Ser Ser
35 965 970 975

Ser Ile Gln Met Ala Met
980

40 <210> 3

<211> 4188

45 <212> DNA

<213> Homo sapiens

<300>

50 <308> genbank

<309> 2002-10-04

55 <313> (1)...(4188)

<300>
 <308> genbank NM_016281
 5 <309> 2002-10-04
 <313> (1)..(4188)
 <400> 3
 10 gccgggaaac aagccacagg agagcgactc aggaacaagt gtgggagagg aagcggcggc 60
 ggcggcgccg ggcccgaaaa tggtgacagc aggtctgagg ttgcataata aatacaaagg 120
 15 actgaagtta taaaagagaa aagagaagtt tgctgctaaa atgaatctga gcaatatgga 180
 atatttgtg ccacacacaa aaaggtaactg aagatttacc ccccaaaaaa aattgtcaat 240
 20 gagaataaa gctaactgat atcaaaaagc agagcctgct ctactggcca tcatgcgtaa 300
 aggggtgctg aaggacccag agattgccga tctattctac aaagatgatc ctgaggaact 360
 ttttatttgtt ttgcataaaa ttggacatgg aagtttggaa gcagtttatt ttgcataaaa 420
 25 tgctcacacc aatgaggtgg tggcaattaa gaagatgtcc tatagtggaa agcagaccca 480
 tgagaaatgg caagatattc ttaaggaagt taaatttttca cgacaattga agcatcctaa 540
 tactattttagg tacaaaggct gttacttgaa agaacacact gcttggtgg tcatggataa 600
 30 ttgccttaggc tcagcctctg atttattttaga agttcataaaa aaaccacttc aggaagtggaa 660
 gatcgctgcc attactcatg gacccctgca tggacttagcc tacctacatt ctcatgcatt 720
 35 gattcatagg gatattaaag cagggaaatat tcttctaaca gagccaggc agttaaaact 780
 agctgatttt ggatctgctt caatggcttc tcctgccaac tccttcgtgg gcacacctta 840
 ctggatggct ccagaggtga tcttagctat ggatgaagga cagttatgtg ggaaagttga 900
 40 tatttttgtca cttggcatca cttgtattga attggcgaa cggaaagccgc ccctttcaa 960
 catgaatgca atgagtgccct tatatcacat tgccccagaat gactccccaa cgttacagtc 1020
 45 taatgaatgg acagactcct ttaggagatt tggtgattac tgcttgccaga aaatacctca 1080
 ggaaaggcca acatcagcag aactattaag gcatgacttt gttcgacgag accggccact 1140
 50 acgtgtccctc attgacccatca tacagaggac aaaagatgca gttcgatgagc tagataaccc 1200
 acagtaccga aaaatgaaaa aaatactttt ccaagagaca cggaaatggac cttgaatga 1260
 gtcacaggag gatgaggaag acagtgaaca tggaaaccagc ctgaacaggaaatggacag 1320
 55 cctgggcagc aaccattcca ttccaagcat gtcgtgagc acaggcagcc agagcagcag 1380
 tgtgaacagc atgcaggaag tcatggacga gaggcgttcc gaacttgtca tcatgcacga 1440

5	tgacgaaagc acaatcaatt ccagctcctc cgtcggtcat aagaaagatc atgtattcat	1500
	aaggatgag gcgggccacg gcgtcccacg gcctgagccg cggcctaccc agtcagttca	1560
	gagccaggcc ctccactacc ggaacagaga gcgcttgcc acgatcaa at cagcatctt	1620
	ggttacacga cagatccatg agcatgagca ggagaacgag ttgcgggaac agatgtcagg	1680
10	ttataagcgg atgcggcgcc agcaccagaa gcagctgatc gccctggaga acaagctgaa	1740
	ggctgagatg gacgagcacc gcctcaagct acagaaggag gtggagacgc atgccaacaa	1800
	ctcgccatc gagctggaga agctggccaa gaagcaagtg gctatcatag aaaaggaggc	1860
15	aaaggtagct gcagcagatg agaagaagtt ccagcaacag atcttggccc agcagaagaa	1920
	agatttgaca actttcttag aaagtcagaa gaagcagtat aagatttgta aggaaaaat	1980
20	aaaagaggaa atgaatgagg accatagcac acccaagaaa gagaagcaag agcggatctc	2040
	caaacataaa gagaacttgc agcacacaca ggctgaagag gaagcccacc ttctcaactca	2100
	acagagactg tactacgaca aaaattgtcg tttcttcaag cgaaaaataa tgatcaagcg	2160
25	gcacgaggtg gagcagcaga acattcgga ggaactaaat aaaaagagga cccagaagga	2220
	gatggagcat gccatgctaa tccggcacga cgagtccacc cgagagctag agtacaggca	2280
30	gctgcacacg ttacagaagc tacgcatgga tctgatccgt ttacagcacc agacggaact	2340
	ggaaaaccag ctggagtaca ataagaggcg agaaagagaa ctgcacagaa agcatgtcat	2400
	ggaacttcgg caacagccaa aaaactaaa ggccatggaa atgcaaatta aaaaacagtt	2460
35	tcaggacact tgcaaagtac agaccaaaca gtataaagca ctcaagaatc accagttgga	2520
	agttactcca aagaatgagc acaaaacaat cttaaagaca ctgaaagatg agcagacaag	2580
40	aaaacttgcc atttggcag agcagtatga acagagtata aatgaaatga tggcctctca	2640
	agcgtaacgg ctagatgagg ctcaagaagc agaatgccag gccttgaggc tacagctcca	2700
	gcagggaaatg gagctgctca acgcctacca gagaaaatc aagatgcaaa cagaggcaca	2760
45	acatgaacgt gagctccaga agctagagca gagagtgtct ctgcgcagag cacaccttga	2820
	gcagaagatt gaagaggagc tggctgcct tcagaaggaa cgcagcaga gaataaaagaa	2880
50	cctattggaa aggcaagagc gagagattga aacttttgc atggagagcc tcagaatggg	2940
	atttggaaat ttggttacat tagatttcc taaggaggac tacagatgag attaaatttt	3000
	ttgcattta caaaaaaaaaaaaaaaaaga aaacagaaaa aaattcagac cctgcaaaac	3060
55	cacattcccc attttaacgg gcgttgctct cactctctct ctctcttact cttactgaca	3120

tcgtgtcgga ctagtgcctg tttattctta ctccatcagg ggcccccttc ctccccccgt 3180
gtcaacttgc agtgctggcc aaaacctggc cgtctcttct attcacagta cacgtcacag 3240
5 tattgatgtg attcaaaatg tttcagtgaa aactttggag acagtttaa caaaaccaat 3300
aaaccaacaa caaaaaaaagt ggtatgtat tgctttaagc aatcactcat taccaccaat 3360
10 ctgtgaaagt aaagcaaaaa ataataataa taaatgccaa gggggagaga gacacaatat 3420
ccgcagcctt acacctaacc tagctgctgc attatttat tttatttat tttttggta 3480
tttattcatc aggaataaaaa aaaacaaagt tttattnaag attgaaaatt tgatacattt 3540
15 tacagaaaact aattgtgatg tacatatacg tggtgacata ttattacttt ttggggacg 3600
ggggtgggtg gggtaagag atcttgat ttttagactgc tgcagagttt acttgcctca 3660
20 gcatatctga tgtatcataa tcattctgc tgtgcagagg agggatacac ttaggggctc 3720
acagatccca gtagcacaat tgggcttgg caaatggta ttttggat agaggaattt 3780
aaggagaggt attacttatt ttcattttgt attttaactg tttctcgat caaatttttt 3840
25 aacttcttc tcgtgttctt ccccacctcc ttccctttcc agttcagtat ttggagttca 3900
acactgtctc tcaatcagat catctggatc ttttcttta tctcccttcc cttccctaag 3960
30 tcccatttct tggtcataaa tattgcatta ttcacacttt caaactgtgt attttcttac 4020
aataaaaaat gatgaaaaaaa aaaaaggctt tacttcttt gcatgcactt taaaaacaaa 4080
acaacacatt tttcagggttc caaggaagag catgataact gtcagagctt ttaatttata 4140
35 ttgtaaataa aagtgttcat cacaaaaaaaaaaaaaaaaaaaaaaaa 4188

40 <210> 4
40 <211> 898
45 <212> PRT
<213> Homo sapiens
<300>
50 <308> Genbank NP_057365.2
<309> 2002-10-01
<313> (1)..(898)
55 <400> 4

Met Arg Lys Gly Val Leu Lys Asp Pro Glu Ile Ala Asp Leu Phe Tyr
 1 5 10 15

Lys Asp Asp Pro Glu Glu Leu Phe Ile Gly Leu His Glu Ile Gly His
 5 20 25 30

Gly Ser Phe Gly Ala Val Tyr Phe Ala Thr Asn Ala His Thr Asn Glu
 35 40 45

10 Val Val Ala Ile Lys Lys Met Ser Tyr Ser Gly Lys Gln Thr His Glu
 50 55 60

Lys Trp Gln Asp Ile Leu Lys Glu Val Lys Phe Leu Arg Gln Leu Lys
 65 70 75 80

15 His Pro Asn Thr Ile Glu Tyr Lys Gly Cys Tyr Leu Lys Glu His Thr
 85 90 95

Ala Trp Leu Val Met Glu Tyr Cys Leu Gly Ser Ala Ser Asp Leu Leu
 20 100 105 110

Glu Val His Lys Lys Pro Leu Gln Glu Val Glu Ile Ala Ala Ile Thr
 115 120 125

25 His Gly Ala Leu His Gly Leu Ala Tyr Leu His Ser His Ala Leu Ile
 130 135 140

His Arg Asp Ile Lys Ala Gly Asn Ile Leu Leu Thr Glu Pro Gly Gln
 145 150 155 160

30 Val Lys Leu Ala Asp Phe Gly Ser Ala Ser Met Ala Ser Pro Ala Asn
 165 170 175

Ser Phe Val Gly Thr Pro Tyr Trp Met Ala Pro Glu Val Ile Leu Ala
 35 180 185 190

Met Asp Glu Gly Gln Tyr Asp Gly Lys Val Asp Ile Trp Ser Leu Gly
 195 200 205

40 Ile Thr Cys Ile Glu Leu Ala Glu Arg Lys Pro Pro Leu Phe Asn Met
 210 215 220

Asn Ala Met Ser Ala Leu Tyr His Ile Ala Gln Asn Asp Ser Pro Thr
 225 230 235 240

45 Leu Gln Ser Asn Glu Trp Thr Asp Ser Phe Arg Arg Phe Val Asp Tyr
 245 250 255

Cys Leu Gln Lys Ile Pro Gln Glu Arg Pro Thr Ser Ala Glu Leu Leu
 50 260 265 270

Arg His Asp Phe Val Arg Arg Asp Arg Pro Leu Arg Val Leu Ile Asp
 275 280 285

55 Leu Ile Gln Arg Thr Lys Asp Ala Val Arg Glu Leu Asp Asn Leu Gln
 290 295 300

Tyr Arg Lys Met Lys Lys Ile Leu Phe Gln Glu Thr Arg Asn Gly Pro
 305 310 315 320
 Leu Asn Glu Ser Gln Glu Asp Glu Asp Ser Glu His Gly Thr Ser
 5 325 330 335
 Leu Asn Arg Glu Met Asp Ser Leu Gly Ser Asn His Ser Ile Pro Ser
 340 345 350
 10 Met Ser Val Ser Thr Gly Ser Gln Ser Ser Ser Val Asn Ser Met Gln
 355 360 365
 Glu Val Met Asp Glu Ser Ser Ser Glu Leu Val Met Met His Asp Asp
 370 375 380
 15 Glu Ser Thr Ile Asn Ser Ser Ser Val Val His Lys Lys Asp His
 385 390 395 400
 Val Phe Ile Arg Asp Glu Ala Gly His Gly Asp Pro Arg Pro Glu Pro
 20 405 410 415
 Arg Pro Thr Gln Ser Val Gln Ser Gln Ala Leu His Tyr Arg Asn Arg
 420 425 430
 25 Glu Arg Phe Ala Thr Ile Lys Ser Ala Ser Leu Val Thr Arg Gln Ile
 435 440 445
 His Glu His Glu Gln Glu Asn Glu Leu Arg Glu Gln Met Ser Gly Tyr
 450 455 460
 30 Lys Arg Met Arg Arg Gln His Gln Lys Gln Leu Ile Ala Leu Glu Asn
 465 470 475 480
 Lys Leu Lys Ala Glu Met Asp Glu His Arg Leu Lys Leu Gln Lys Glu
 35 485 490 495
 Val Glu Thr His Ala Asn Asn Ser Ser Ile Glu Leu Glu Lys Leu Ala
 500 505 510
 40 Lys Lys Gln Val Ala Ile Ile Glu Lys Glu Ala Lys Val Ala Ala Ala
 515 520 525
 Asp Glu Lys Lys Phe Gln Gln Ile Leu Ala Gln Gln Lys Lys Asp
 530 535 540
 45 Leu Thr Thr Phe Leu Glu Ser Gln Lys Lys Gln Tyr Lys Ile Cys Lys
 545 550 555 560
 Glu Lys Ile Lys Glu Glu Met Asn Glu Asp His Ser Thr Pro Lys Lys
 50 565 570 575
 Glu Lys Gln Glu Arg Ile Ser Lys His Lys Glu Asn Leu Gln His Thr
 580 585 590
 55 Gln Ala Glu Glu Glu Ala His Leu Leu Thr Gln Gln Arg Leu Tyr Tyr
 595 600 605

Asp Lys Asn Cys Arg Phe Phe Lys Arg Lys Ile Met Ile Lys Arg His
 610 615 620

5 Glu Val Glu Gln Gln Asn Ile Arg Glu Glu Leu Asn Lys Lys Arg Thr
 625 630 635 640

Gln Lys Glu Met Glu His Ala Met Leu Ile Arg His Asp Glu Ser Thr
 645 650 655

10 Arg Glu Leu Glu Tyr Arg Gln Leu His Thr Leu Gln Lys Leu Arg Met
 660 665 670

Asp Leu Ile Arg Leu Gln His Gln Thr Glu Leu Glu Asn Gln Leu Glu
 15 675 680 685

Tyr Asn Lys Arg Arg Glu Arg Glu Leu His Arg Lys His Val Met Glu
 690 695 700

20 Leu Arg Gln Gln Pro Lys Asn Leu Lys Ala Met Glu Met Gln Ile Lys
 705 710 715 720

Lys Gln Phe Gln Asp Thr Cys Lys Val Gln Thr Lys Gln Tyr Lys Ala
 725 730 735

25 Leu Lys Asn His Gln Leu Glu Val Thr Pro Lys Asn Glu His Lys Thr
 740 745 750

Ile Leu Lys Thr Leu Lys Asp Glu Gln Thr Arg Lys Leu Ala Ile Leu
 30 755 760 765

Ala Glu Gln Tyr Glu Gln Ser Ile Asn Glu Met Met Ala Ser Gln Ala
 770 775 780

35 Leu Arg Leu Asp Glu Ala Gln Glu Ala Glu Cys Gln Ala Leu Arg Leu
 785 790 795 800

Gln Leu Gln Gln Glu Met Glu Leu Leu Asn Ala Tyr Gln Ser Lys Ile
 805 810 815

40 Lys Met Gln Thr Glu Ala Gln His Glu Arg Glu Leu Gln Lys Leu Glu
 820 825 830

Gln Arg Val Ser Leu Arg Arg Ala His Leu Glu Gln Lys Ile Glu Glu
 45 835 840 845

Glu Leu Ala Ala Leu Gln Lys Glu Arg Ser Glu Arg Ile Lys Asn Leu
 850 855 860

50 Leu Glu Arg Gln Glu Arg Glu Ile Glu Thr Phe Asp Met Glu Ser Leu
 865 870 875 880

Arg Met Gly Phe Gly Asn Leu Val Thr Leu Asp Phe Pro Lys Glu Asp
 885 890 895

55 Tyr Arg

5 <210> 5
 5 <211> 4971
 5 <212> DNA
 5 <213> Homo sapiens
 10 <300>
 10 <308> Genbank NM_016151.1
 15 <309> 2002-11-05
 15 <313> (1)..(4971)
 20 <400> 5
 20 aattcggcac gagctgagac ggagaagagg agagggcagag agggcgcggg gaccgtcagc 60
 20 agcaccttag ctacaatcg tcaagctattc tcggaagaga gaagggagag ggaggaggcc 120
 25 ggggcgggag tgggggctgt caccctcggc ccccgccgtg agaggggccc tgccggccgg 180
 25 cgtcctcggg gtggggccccc agtcggtgcc cgaagaccta cagctcaggc ccctgggtcc 240
 30 caaatttcca ggcttgccc ctccctcctt ctcagatacc cggtaacag tcctcatagt 300
 30 ccagatatcc gggactcggg tcccaacctc tctaaacctg ggtctctgtt tcatagatt 360
 35 tcaaataatca ggttcaggcc cctgcgtgca ccagtatccg gggttcattc cccgggcgtt 420
 35 tcaaataatcg gattcagtttccatccgtt cagatattcg gggttcagac cccacaatca 480
 40 gaaatccgga attcggcagc tgtcgcctc gacgaggggg aggactggac cgcgaggtca 540
 40 gattaggttgc tcaaaaaatccctc ccctccaggg gaggcttccc gggcccgccc ctcaggaagg 600
 45 gcgaaagccg aggaagaggt ggcaaggggg aaggctcctc tgccctctc cctgcttggc 660
 45 agagccgctg gaggacccca ggcggaaagcg gaggcgctgg ggcaccatag tgaccctac 720
 50 caggccaggg cccactctca gggcccccag gggccaccat gccagctggg ggccggccgg 780
 50 ggagccctgaa ggaccaggat gtggctgagc tcttcttcaa ggatgacccca gaaaagctct 840
 55 tctctgacccctc ccgggaaatt ggccatggca gctttggagc cgtataacttt gcccggatg 900
 55 tccggaatag tgaggtggtg gccatcaaga agatgtccta cagtggaaag cagtccatg 960
 55 agaaatggca agacatcatc aaggagggtgc gtttcttaca gaagctccgg catcccaaca 1020
 55 ccattcagta ccggggctgt tacctgaggg agcacacggc ttggctggta atggagtatt 1080
 55 gcctgggctc agttctgac cttctagaag tgcacaagaa accccttcag gagtagaga 1140

5	tcgcagctgt gacccacggg gcgccagg gcctggcata tctgcactcc cacaacatga	1200
	tccatagggta tgtgaaggct ggaaacatcc tgctgtcaga gccagggtta gtgaagctag	1260
	gggactttgg ttctgcgtcc atcatggcac ctgccaactc cttcgtgggc accccatact	1320
	ggatggcacc cgaggtgatc ctggccatgg atgaggggca gtacgatggc aaagtggacg	1380
10	tctggtcctt gggataacc tgcacatcgagc tggctgaacg gaaaccacccg ctcttaaca	1440
	tgaatgcgat gagtgccctta taccacattt cacagaacga atccccgtg ctccagtcag	1500
	gacactggtc tgagtacttc cgaaattttg tcgactcctg tcttcagaaa atccctcaag	1560
15	acagaccaac ctcagaggtt ctccctgaagc accgctttgt gctccggag cggccaccca	1620
	cagtcatcat ggacctgatc cagaggacca aggatgccgt gcgggagctg gacaacctgc	1680
20	agtaccgcaa gatgaagaag atcctgttcc aagaggcacc caacggccct ggtgccgagg	1740
	ccccagagga ggaagaggag gccgagccct acatgcacccg ggccggact ctgaccagcc	1800
	tcgagagtag ccactcagtg cccagcatgt ccatcagcgc ctccagccag agcagctccg	1860
25	tcaacagcct agcagatgcc tcagacaacg aggaagagga ggaggaggag gaggaagagg	1920
	aggaggagga agaaggccct gaagccccc agatggccat gatgcaggag ggggagcaca	1980
30	cagtcaccc tcacagctcc attatccacc ggctgccggg ctctgacaac ctatatgtat	2040
	acccttacca gccagagata acccccagcc ctctccagcc gcctgcagcc ccagctccca	2100
	cttccaccac ctttccggc cgccgccccg cctactgccg taaccgagac cactttgcca	2160
35	ccatccgaac cgccctccctg gtcagccgtc agatccagga gcatgagcag gactctgcgc	2220
	tgcgggagca gctgagcggc tataagcggta tgcgacgaca gcaccagaag cagctgctgg	2280
40	ccctggagtc acggctgagg ggtgaacggg aggagcacag tgcacggctg cagcggagc	2340
	ttgaggcgca gcgggctggc tttggggcag aggcagaaaa gctggcccg cgccaccagg	2400
	ccataggtga gaaggaggca cgagctgccc aggccgagga gcggaaatgc cagcagcaca	2460
45	tccttggca gcagaagaag gagctggctg ccctgctgga ggcacagaag cggacctaca	2520
	aacttcgcaa ggaacagctg aaggaggagc tccaggagaa ccccagcaact cccaaagcggg	2580
50	agaaggccga gtggctgctg cggcagaagg agcagctcca gcagtgcacag gcggaggagg	2640
	aagcagggct gctgcggcgg cagcgccagt actttgagct gcagtgtcgc cagtacaagc	2700
	gcaagatgtt gctggctcgg cacagcctgg accaggaccc gctgcggag gacctgaaca	2760
55	agaagcagac ccagaaggac ttggagtgtg cactgctgct tcggcagcac gaggccacgc	2820

gggagctgga	gctgcggcag	ctccaggccg	tgcagcgcac	gcgggctgag	ctcacccgcc	2880	
5	tgcagcacca	gacggagctg	ggcaaccagc	tggagtacaa	caagcggcgt	gagcaagagt	2940
10	tgcggcagaa	gcatgcggcc	cagttcgcc	agcagccaa	gagcctcaa	gtacgtgcag	3000
15	gccagcggcc	cccgccctt	ccactcccc	tccctgggc	tctggccca	ccaaacacag	3060
20	gcacccctat	agaacagcag	ccctgctcac	ctggccagga	ggcagtccctg	gaccaaagaa	3120
25	tgcttggcga	ggaggagaa	gcagttggag	agagaaggat	tctggaaag	gaagggggca	3180
30	ctttggagcc	caagcagcag	aggattctgg	gggaagaatc	aggagccct	agtcccagtc	3240
35	cacaaaaaca	tgggagcctg	gttcatgagg	aagttgggg	tctgcctgag	gagatagagg	3300
40	agcttagggt	gccctccctt	gtaccccagg	agaggagcat	tgttggccag	gaggaggctg	3360
45	ggacgtggag	cttgggggg	aaggaggatg	agagtcttct	ggatgaggag	ttttagctt	3420
50	gctgggtcca	gggcccagca	ctgactcccg	tccctgagga	ggaggaagaa	gaggaagagg	3480
55	gggctccgat	tgggaccctt	aggatccctg	gagatggttt	tccttcccc	gacatccctc	3540
60	ctgaacccccc	tccaacacac	ctgaggccct	gccctgccag	ccagctccct	ggactcctgt	3600
65	cccatggccct	cctggccggc	ctctcctttt	cagtgggttc	ctcctctggc	ctcctgcccc	3660
70	tcctgctgct	gctgctgctt	ccattgctgg	cagcccaggg	tgggggtggc	ctgcaggcag	3720
75	cgctgctggc	ccttgaggtg	gggctgggtgg	gtctgggggc	ctcctacctg	ctcctttgtt	3780
80	cagccctgca	cctgcccctcc	agtctttcc	tactcctggc	ccagggtacc	gcactgggg	3840
85	ccgtcctggg	cctgagctgg	cgccgaggcc	tcatgggtgt	tcccctggc	cttggagctg	3900
90	cctggctctt	agcttggcca	ggcctagctc	tacctctggt	ggctatggca	gcggggggca	3960
95	gatgggtgcg	gcagcagggc	ccccgggtgc	gccggggcat	atctcgactc	tggttgcggg	4020
100	ttctgctgctg	cctgtcaccc	atggccttcc	gggcctctgca	gggctgtggg	gctgtgggg	4080
105	accgggtct	gtttgcactg	tacccaaaaa	ccaacaagga	tggctccgc	agccgcctgc	4140
110	ccgtccctgg	gccccggccg	cgtaatcccc	gcaccaccca	acacccatta	gctctgttgg	4200
115	caagggctcg	ggtcctgtgc	aagggctgga	actggcgtct	ggcacgggcc	agccagggtt	4260
120	tagcatccca	cttgcccccg	tggccatcc	acacactggc	cagctggggc	ctgcttcggg	4320
125	gtgaacggcc	cacccgaatc	ccccggctac	taccacgcag	ccagcggccag	ctagggcccc	4380
130	ctgcctccca	ccagccactg	ccagggactc	tagccggccg	gaggtcacgc	acccgcccagt	4440
135	cccgccct	gccccctgg	aggtagctga	ctccagccct	tccagccaa	atctagagca	4500
140	ttgagcactt	tatctccac	gactcagtga	agtttctcca	gtccctagtc	ctctcttttc	4560

acccaccttc ctcagttgc tcacttaccc caggcccagc ccttcggacc tctagacagg 4620
 cagcctcctc agctgtggag tccagcagtc actctgtgtt ctccctggcgc tcctccccta 4680
 5 agttattgct gttcgccgc tgtgtgtgct catcctcacc ctcattgact caggcctggg 4740
 gccaggggtg gtggaggggtg ggaagagtca tgttttttt ctccctttt attttgtttt 4800
 10 tctgtctccc ttccaaacctg tcccccttccc cccacccaaa aaagaaaaag acaaacacaa 4860
 ataaaaatatc tgagcggAAC tgtgaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4920
 15 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 4971
 <210> 6
 20 <211> 1235
 <212> PRT
 <213> Homo sapiens
 25 <300>
 <308> Genbank NP_057235.1
 30 <309> 2002-11-05
 <313> (1)..(1235)
 35 <400> 6
 Met Pro Ala Gly Gly Arg Ala Gly Ser Leu Lys Asp Pro Asp Val Ala
 1 5 10 15
 40 Glu Leu Phe Phe Lys Asp Asp Pro Glu Lys Leu Phe Ser Asp Leu Arg
 20 25 30
 Glu Ile Gly His Gly Ser Phe Gly Ala Val Tyr Phe Ala Arg Asp Val
 35 40 45
 45 Arg Asn Ser Glu Val Val Ala Ile Lys Lys Met Ser Tyr Ser Gly Lys
 50 55 60
 Gln Ser Asn Glu Lys Trp Gln Asp Ile Ile Lys Glu Val Arg Phe Leu
 50 65 70 75 80
 Gln Lys Leu Arg His Pro Asn Thr Ile Gln Tyr Arg Gly Cys Tyr Leu
 85 90 95
 55 Arg Glu His Thr Ala Trp Leu Val Met Glu Tyr Cys Leu Gly Ser Ala
 100 105 110

Ser Asp Leu Leu Glu Val His Lys Lys Pro Leu Gln Glu Val Glu Ile
115 120 125

5 Ala Ala Val Thr His Gly Ala Leu Gln Gly Leu Ala Tyr Leu His Ser
130 135 140

His Asn Met Ile His Arg Asp Val Lys Ala Gly Asn Ile Leu Leu Ser
145 150 155 160

10 Glu Pro Gly Leu Val Lys Leu Gly Asp Phe Gly Ser Ala Ser Ile Met
165 170 175

Ala Pro Ala Asn Ser Phe Val Gly Thr Pro Tyr Trp Met Ala Pro Glu
180 185 190

15 Val Ile Leu Ala Met Asp Glu Gly Gln Tyr Asp Gly Lys Val Asp Val
195 200 205

Trp Ser Leu Gly Ile Thr Cys Ile Glu Leu Ala Glu Arg Lys Pro Pro
210 215 220

20 Leu Phe Asn Met Asn Ala Met Ser Ala Leu Tyr His Ile Ala Gln Asn
225 230 235 240

25 Glu Ser Pro Val Leu Gln Ser Gly His Trp Ser Glu Tyr Phe Arg Asn
245 250 255

Phe Val Asp Ser Cys Leu Gln Lys Ile Pro Gln Asp Arg Pro Thr Ser
260 265 270

30 Glu Val Leu Leu Lys His Arg Phe Val Leu Arg Glu Arg Pro Pro Thr
275 280 285

35 Val Ile Met Asp Leu Ile Gln Arg Thr Lys Asp Ala Val Arg Glu Leu
290 295 300

Asp Asn Leu Gln Tyr Arg Lys Met Lys Lys Ile Leu Phe Gln Glu Ala
305 310 315 320

40 Pro Asn Gly Pro Gly Ala Glu Ala Pro Glu Glu Glu Glu Ala Glu
325 330 335

Pro Tyr Met His Arg Ala Gly Thr Leu Thr Ser Leu Glu Ser Ser His
340 345 350

45 Ser Val Pro Ser Met Ser Ile Ser Ala Ser Ser Gln Ser Ser Ser Val
355 360 365

Asn Ser Leu Ala Asp Ala Ser Asp Asn Glu Glu Glu Glu Glu Glu
50 370 375 380

Glu Glu Glu Glu Glu Glu Glu Gly Pro Glu Ala Arg Glu Met Ala
385 390 395 400

55 Met Met Gln Glu Gly Glu His Thr Val Thr Ser His Ser Ser Ile Ile
405 410 415

	His	Arg	Leu	Pro	Gly	Ser	Asp	Asn	Leu	Tyr	Asp	Asp	Pro	Tyr	Gln	Pro	
				420					425				430				
5	Glu	Ile	Thr	Pro	Ser	Pro	Leu	Gln	Pro	Pro	Ala	Ala	Pro	Ala	Pro	Thr	
				435				440				445					
	Ser	Thr	Thr	Ser	Ser	Ala	Arg	Arg	Arg	Arg	Ala	Tyr	Cys	Arg	Asn	Arg	Asp
				450			455				460						
10	His	Phe	Ala	Thr	Ile	Arg	Thr	Ala	Ser	Leu	Val	Ser	Arg	Gln	Ile	Gln	
				465			470			475			480				
	Glu	His	Glu	Gln	Asp	Ser	Ala	Leu	Arg	Glu	Gln	Leu	Ser	Gly	Tyr	Lys	
15					485				490			495					
	Arg	Met	Arg	Arg	Gln	His	Gln	Lys	Gln	Leu	Leu	Ala	Leu	Glu	Ser	Arg	
				500				505			510						
20	Leu	Arg	Gly	Glu	Arg	Glu	Glu	His	Ser	Ala	Arg	Leu	Gln	Arg	Glu	Leu	
				515			520				525						
	Glu	Ala	Gln	Arg	Ala	Gly	Phe	Gly	Ala	Glu	Ala	Glu	Lys	Leu	Ala	Arg	
		530				535			540								
25	Arg	His	Gln	Ala	Ile	Gly	Glu	Lys	Glu	Ala	Arg	Ala	Ala	Gln	Ala	Glu	
		545				550			555			560					
	Glu	Arg	Lys	Phe	Gln	Gln	His	Ile	Leu	Gly	Gln	Gln	Lys	Glu	Leu		
				565			570			575							
30	Ala	Ala	Leu	Leu	Glu	Ala	Gln	Lys	Arg	Thr	Tyr	Lys	Leu	Arg	Lys	Glu	
				580			585			590							
	Gln	Leu	Lys	Glu	Glu	Leu	Gln	Glu	Asn	Pro	Ser	Thr	Pro	Lys	Arg	Glu	
35			595			600			605								
	Lys	Ala	Glu	Trp	Leu	Leu	Arg	Gln	Lys	Glu	Gln	Leu	Gln	Gln	Cys	Gln	
		610			615				620								
40	Ala	Glu	Glu	Ala	Gly	Leu	Leu	Arg	Arg	Gln	Arg	Gln	Tyr	Phe	Glu		
		625			630			635			640						
	Leu	Gln	Cys	Arg	Gln	Tyr	Lys	Arg	Lys	Met	Leu	Leu	Ala	Arg	His	Ser	
			645			650			655								
45	Leu	Asp	Gln	Asp	Leu	Leu	Arg	Glu	Asp	Leu	Asn	Lys	Lys	Gln	Thr	Gln	
		660			665			670									
	Lys	Asp	Leu	Glu	Cys	Ala	Leu	Leu	Arg	Gln	His	Glu	Ala	Thr	Arg		
50			675			680			685								
	Glu	Leu	Glu	Leu	Arg	Gln	Leu	Gln	Ala	Val	Gln	Arg	Thr	Arg	Ala	Glu	
		690			695			700									
55	Leu	Thr	Arg	Leu	Gln	His	Gln	Thr	Glu	Leu	Gly	Asn	Gln	Leu	Glu	Tyr	
		705			710			715			720						

Asn Lys Arg Arg Glu Gln Glu Leu Arg Gln Lys His Ala Ala Gln Val
 725 730 735

5 Arg Gln Gln Pro Lys Ser Leu Lys Val Arg Ala Gly Gln Arg Pro Pro
 740 745 750

Gly Leu Pro Leu Pro Ile Pro Gly Ala Leu Gly Pro Pro Asn Thr Gly
 755 760 765

10 Thr Pro Ile Glu Gln Gln Pro Cys Ser Pro Gly Gln Glu Ala Val Leu
 770 775 780

Asp Gln Arg Met Leu Gly Glu Glu Glu Ala Val Gly Glu Arg Arg
 785 790 795 800

15 Ile Leu Gly Lys Glu Gly Ala Thr Leu Glu Pro Lys Gln Gln Arg Ile
 805 810 815

20 Leu Gly Glu Glu Ser Gly Ala Pro Ser Pro Ser Pro Gln Lys His Gly
 820 825 830

Ser Leu Val Asp Glu Glu Val Trp Gly Leu Pro Glu Glu Ile Glu Glu
 835 840 845

25 Leu Arg Val Pro Ser Leu Val Pro Gln Glu Arg Ser Ile Val Gly Gln
 850 855 860

Glu Glu Ala Gly Thr Trp Ser Leu Trp Gly Lys Glu Asp Glu Ser Leu
 865 870 875 880

30 Leu Asp Glu Glu Phe Glu Leu Gly Trp Val Gln Gly Pro Ala Leu Thr
 885 890 895

35 Pro Val Pro Glu Glu Glu Glu Glu Glu Glu Gly Ala Pro Ile Gly
 900 905 910

Thr Pro Arg Asp Pro Gly Asp Gly Cys Pro Ser Pro Asp Ile Pro Pro
 915 920 925

40 Glu Pro Pro Pro Thr His Leu Arg Pro Cys Pro Ala Ser Gln Leu Pro
 930 935 940

Gly Leu Leu Ser His Gly Leu Leu Ala Gly Leu Ser Phe Ala Val Gly
 945 950 955 960

45 Ser Ser Ser Gly Leu Leu Pro Leu Leu Leu Leu Leu Leu Pro Leu
 965 970 975

50 Leu Ala Ala Gln Gly Gly Leu Gln Ala Ala Leu Leu Ala Leu
 980 985 990

Glu Val Gly Leu Val Gly Leu Gly Ala Ser Tyr Leu Leu Leu Cys Thr
 995 1000 1005

55 Ala Leu His Leu Pro Ser Ser Leu Phe Leu Leu Ala Gln Gly
 1010 1015 1020

Thr Ala Leu Gly Ala Val Leu Gly Leu Ser Trp Arg Arg Gly Leu
 1025 1030 1035

5 Met Gly Val Pro Leu Gly Leu Gly Ala Ala Trp Leu Leu Ala Trp
 1040 1045 1050

Pro Gly Leu Ala Leu Pro Leu Val Ala Met Ala Ala Gly Gly Arg
 1055 1060 1065

10 Trp Val Arg Gln Gln Gly Pro Arg Val Arg Arg Gly Ile Ser Arg
 1070 1075 1080

Leu Trp Leu Arg Val Leu Leu Arg Leu Ser Pro Met Ala Phe Arg
 1085 1090 1095

15 Ala Leu Gln Gly Cys Gly Ala Val Gly Asp Arg Gly Leu Phe Ala
 1100 1105 1110

20 Leu Tyr Pro Lys Thr Asn Lys Asp Gly Phe Arg Ser Arg Leu Pro
 1115 1120 1125

Val Pro Gly Pro Arg Arg Arg Asn Pro Arg Thr Thr Gln His Pro
 1130 1135 1140

25 Leu Ala Leu Leu Ala Arg Val Trp Val Leu Cys Lys Gly Trp Asn
 1145 1150 1155

Trp Arg Leu Ala Arg Ala Ser Gln Gly Leu Ala Ser His Leu Pro
 1160 1165 1170

30 Pro Trp Ala Ile His Thr Leu Ala Ser Trp Gly Leu Leu Arg Gly
 1175 1180 1185

Glu Arg Pro Thr Arg Ile Pro Arg Leu Leu Pro Arg Ser Gln Arg
 35 1190 1195 1200

Gln Leu Gly Pro Pro Ala Ser His Gln Pro Leu Pro Gly Thr Leu
 1205 1210 1215

40 Ala Gly Arg Arg Ser Arg Thr Arg Gln Ser Arg Ala Leu Pro Pro
 1220 1225 1230

Trp Arg
 45 1235

<210> 7

<211> 4242

50 <212> DNA

<213> Homo sapiens

55 <300>

<308> Genbank NM_004783

<309> 2000-11-01

5 <313> (1)...(4242)

10	<400> 7 agaatttcaa atatcaggtt caggccccctg cgtgcaccag tatccggggt tcattcccg ggcgttcaaa tatcgatc agtctccatc ccgttcagat attcggggtt cagacccac aatcagaaat ccggaattcg gcagctgtcg ccctcgacga gggggaggac tggaccgcga 15 ggtcagatta ggttgcacc ccctccccctc cagggggagggc ttcccccggcc cgccccctcag gaagggcgaa agccgagggaa gaggtggcaa gggaaaggt ctccctgccc ctctccctgc 20 ttggcagagc cgctggagga ccccaggcgg aagcggagggc gctggggcac catagtgacc cctaccaggc caggccccac tctcaggggcc cccagggggcc accatgccag ctggggggccg ggccggggagc ctgaaggacc cagatgtggc tgagctttc ttcaaggatg acccagaaaa 25 gctcttctct gacctccggg aaattggcca tggcagctt ggagccgtat actttgccc ggatgtccgg aatagtgagg tggtgccat caagaagatg tcctacagtg ggaagcagtc 30 caatgagaaa tggcaagaca tcatcaagga ggtgcgggatc ttacagaagc tccggcatcc caacaccatt cagtaccggg gctgttaccc gagggagcac acggcttggc tggtaatgga gtattgcctg ggctcagctt ctgaccttctc agaagtgcac aagaaacccc ttcaggaggt 35 agagatcgca gctgtgaccc acggggcgct tcagggcctg gcatatctgc actccccacaa catgatccat agggatgtga aggctggaaa catcctgctg tcagagccag gtttagtgaa 40 gctaggggac tttgggtctg cgtccatcat ggcacctgcc aactccctcg tggcacc atactggatg gcacccgagg tgatcctggc catggatgag gggcagtacg atggcaaagt ggacgtctgg tccttgggaa taacctgcat cgagctggct gaacggaaac caccgcttct 45 taacatgaat gcgtgatgtg ctttataccca cattgcacag aacgaatccc ccgtgctcca gtcaggacac tggtctgagt acttccggaa ttttgcac tcctgtcttc agaaaatccc 50 tcaagacaga ccaacccatcag aggttctccct gaagcaccgc tttgtgctcc gggagcggcc acccacagtc atcatggacc tgatccagag gaccaaggat gccgtgcggg agctggacaa cctgcagtac cgcaagatga agaagatccct gttccaagag gcacccaaacg gcccctggtgc 55 cgaggccccca gaggaggaag aggaggccga gccctacatg caccggcccg ggactctgac cagcctcgag agtagccact cagtccccag catgtccatc agcgccctcca gccagagcag 1500
----	---

5	ctccgtcaac agcctagcag atgcctcaga caacgaggaa gaggaggagg aggaggaggaa	1560
	agaggaggag gaggaagaag gccctgaagc ccgggagatg gccatgtatgc aggagggggaa	1620
	gcacacagtc acctctcaca gctccattat ccaccggctg ccgggctctg acaacctata	1680
	tgatgacccc taccagccag agataacccc cagccctctc cagccgcctg cagccccagc	1740
10	tcccacttcc accaccttctt ctgcccgcgc ccgggcctac tgccgtaacc gagaccactt	1800
	tgccaccatc cgaaccgcct ccctggtcag ccgtcagatc caggagcatg agcaggactc	1860
	tgcgctgcgg gagcagctga gcggctataa gcggatgcga cgacagcacc agaagcagct	1920
15	gctggccctg gagtcacggc tgaggggtga acgggaggag cacagtgcac ggctgcagcg	1980
	ggagcttgag ggcgcagcggg ctggcttgg ggcagaggca gaaaagctgg cccggcggca	2040
20	ccagggccata ggtgagaagg aggacacgagc tgcccaggcc gaggagcggaa agttccagca	2100
	gcacatcctt gggcagcaga agaaggagct ggctgcctg ctggaggcac agaagcggac	2160
	ctacaaactt cgcaaggaac agctgaagga ggagctccag gagaacccca gcactccaa	2220
25	gcgggagaag gccgagtggc tgctgcggca gaaggagcag ctccagcagt gcccaggcggaa	2280
	ggaggaagca gggctgctgc ggcggcagcg ccagtacttt gagctgcagt gtcgcccagta	2340
30	caagcgcaag atgttgctgg ctccggcacag cctggaccag gacctgctgc gggaggaccc	2400
	gaacaagaag cagacccaga aggacttggaa gtgtgcactg ctgctcggc agcacgaggc	2460
	cacgcgggag ctggagctgc ggcagctcca ggccgtgcag cgacgcggg ctgagctcac	2520
35	ccgcctgcag caccagacgg agctggccaa ccagctggag tacaacaagc ggcgtgagca	2580
	agagttgcgg cagaagcatg cggcccaggt tcgcccagcag cccaaagcc tcaaatactaa	2640
40	ggagctgcag atcaagaagc agttccagga gacgtgtaaatccagactc ggcagtacaa	2700
	ggctctgcga gcacacttgc tggagaccac gcccggactc cagcacaaga gcctccttaa	2760
	gcggctcaag gaagagcaga cccgcaagct ggcatcttg gcggagcagt atgaccagtc	2820
45	catctcagag atgctcagct cacaggcgct gcggcttgat gagaccagg aggcagagtt	2880
	ccaggccctt cggcagcagc ttcaacagga gctggagctg ctcaacgctt accagagcaa	2940
50	gatcaagatc cgcacagaga gccagcacga gagggagctg cgggagctgg agcagagggt	3000
	cgcgcgtgcgg cgggcactgc tggagcagcg ggtggaaagag gagctgctgg ccctgcagac	3060
	aggacgctcc gagcgaatcc gcagtcgtct tgagcggcag gcccgtgaga tcgaggcctt	3120
55	cgatgcggaa agcatgaggc tgggcttctc cagcatggct ctggggggca tccggcgtga	3180

agctgctgcc	cagggctatac	ctgctccacc	ccctgccccca	gcctggccct	cccgccccgt	3240	
5	tcccccgttct	ggggcacact	ggagccatgg	ccctcctcca	ccaggcatgc	cccctccagc	3300
ctggcgtag	ccgtctctgc	tggctccccc	aggccccca	aactggctgg	ggccccccac	3360	
10	acaaaagtggg	acaccccgta	gcggagccct	gctgctgcta	agaaacagcc	cccagcccc	3420
gcggcgggca	gcctcggggg	gcagtggcag	tgagaatgtg	ggccccctg	ctgcccgcgg	3480	
15	gccccgggccc	ctgagccgca	gcaccagtgt	cgcttccac	atcctcaatg	gttcttccca	3540
cttctattcc	tgaggtgcag	cggggaggag	cagatgagct	gggcagggca	gggggtgggtg	3600	
20	gagcctgacc	ctggagggca	ctgagctgga	ggccccctgca	agggtagggg	acaagatgt	3660
ggctccagct	cccctcagac	ctcctcatct	catgagcttc	ttggggctgg	ccagtggccc	3720	
25	agggccagct	tggcgataga	tgcctcaagg	ctgcctggga	gccccgcctc	cctaccatgg	3780
tgccaggggt	ctccctccgc	cacctaggaa	aggagggaga	tgtgcgtgtc	aaatattcat	3840	
30	ctagtcccct	gggggagggg	aagggtgggt	ctagacatac	tatattcaga	gaactatact	3900
accctcacag	tgaggccctc	agacactgcca	cagggcagag	caggtctggg	gcctgaggca	3960	
40	45	gggagaatga	gaggccacct	tactggcagg	aaggatcagg	atggggtctt	4020
tgccctgggtc	tcttcccgta	actgtctgac	gtcctgtgcc	gtcttgcct	ttatctttt	4080	
tttttttttt	taattggat	cagggctggg	gcggggaaac	aagggaagga	ccttggaaagg	4140	
ggctgctccc	aggcctgggg	ggcagtcgtg	ggagccctc	tcaagctgtgg	ggctggcaca	4200	
35	gagccccagg	caagcttta	ataaaactgtt	gttattcta	ac	4242	
<210> 8							
40	<211> 1049						
<212> PRT							
<213> Homo sapiens							
45	<300>						
50	<308> Genbank NP_004774.1						
<309> 2000-11-01							
<313> (1)...(1049)							
55	<400> 8						

Met Pro Ala Gly Gly Arg Ala Gly Ser Leu Lys Asp Pro Asp Val Ala
 1 5 10 15

5 Glu Leu Phe Phe Lys Asp Asp Pro Glu Lys Leu Phe Ser Asp Leu Arg
 20 25 30

Glu Ile Gly His Gly Ser Phe Gly Ala Val Tyr Phe Ala Arg Asp Val
 35 40 45

10 Arg Asn Ser Glu Val Val Ala Ile Lys Lys Met Ser Tyr Ser Gly Lys
 50 55 60

Gln Ser Asn Glu Lys Trp Gln Asp Ile Ile Lys Glu Val Arg Phe Leu
 65 70 75 80

15 Gln Lys Leu Arg His Pro Asn Thr Ile Gln Tyr Arg Gly Cys Tyr Leu
 85 90 95

Arg Glu His Thr Ala Trp Leu Val Met Glu Tyr Cys Leu Gly Ser Ala
 20 100 105 110

Ser Asp Leu Leu Glu Val His Lys Lys Pro Leu Gln Glu Val Glu Ile
 115 120 125

25 Ala Ala Val Thr His Gly Ala Leu Gln Gly Leu Ala Tyr Leu His Ser
 130 135 140

His Asn Met Ile His Arg Asp Val Lys Ala Gly Asn Ile Leu Leu Ser
 145 150 155 160

30 Glu Pro Gly Leu Val Lys Leu Gly Asp Phe Gly Ser Ala Ser Ile Met
 165 170 175

Ala Pro Ala Asn Ser Phe Val Gly Thr Pro Tyr Trp Met Ala Pro Glu
 35 180 185 190

Val Ile Leu Ala Met Asp Glu Gly Gln Tyr Asp Gly Lys Val Asp Val
 195 200 205

40 Trp Ser Leu Gly Ile Thr Cys Ile Glu Leu Ala Glu Arg Lys Pro Pro
 210 215 220

Leu Phe Asn Met Asn Ala Met Ser Ala Leu Tyr His Ile Ala Gln Asn
 45 225 230 235 240

Glu Ser Pro Val Leu Gln Ser Gly His Trp Ser Glu Tyr Phe Arg Asn
 245 250 255

Phe Val Asp Ser Cys Leu Gln Lys Ile Pro Gln Asp Arg Pro Thr Ser
 50 260 265 270

Glu Val Leu Leu Lys His Arg Phe Val Leu Arg Glu Arg Pro Pro Thr
 275 280 285

55 Val Ile Met Asp Leu Ile Gln Arg Thr Lys Asp Ala Val Arg Glu Leu
 290 295 300

Asp Asn Leu Gln Tyr Arg Lys Met Lys Lys Ile Leu Phe Gln Glu Ala
 305 310 315 320

5 Pro Asn Gly Pro Gly Ala Glu Ala Pro Glu Glu Glu Glu Ala Glu
 325 330 335

Pro Tyr Met His Arg Ala Gly Thr Leu Thr Ser Leu Glu Ser Ser His
 340 345 350

10 Ser Val Pro Ser Met Ser Ile Ser Ala Ser Ser Gln Ser Ser Ser Val
 355 360 365

Asn Ser Leu Ala Asp Ala Ser Asp Asn Glu Glu Glu Glu Glu Glu
 370 375 380

15 Glu Glu Glu Glu Glu Glu Gly Pro Glu Ala Arg Glu Met Ala
 385 390 395 400

Met Met Gln Glu Gly Glu His Thr Val Thr Ser His Ser Ser Ile Ile
 20 405 410 415

His Arg Leu Pro Gly Ser Asp Asn Leu Tyr Asp Asp Pro Tyr Gln Pro
 420 425 430

25 Glu Ile Thr Pro Ser Pro Leu Gln Pro Pro Ala Ala Pro Ala Pro Thr
 435 440 445

Ser Thr Thr Ser Ser Ala Arg Arg Arg Ala Tyr Cys Arg Asn Arg Asp
 450 455 460

30 His Phe Ala Thr Ile Arg Thr Ala Ser Leu Val Ser Arg Gln Ile Gln
 465 470 475 480

Glu His Glu Gln Asp Ser Ala Leu Arg Glu Gln Leu Ser Gly Tyr Lys
 35 485 490 495

Arg Met Arg Arg Gln His Gln Lys Gln Leu Leu Ala Leu Glu Ser Arg
 500 505 510

40 Leu Arg Gly Glu Arg Glu Glu His Ser Ala Arg Leu Gln Arg Glu Leu
 515 520 525

Glu Ala Gln Arg Ala Gly Phe Gly Ala Glu Ala Glu Lys Leu Ala Arg
 530 535 540

45 Arg His Gln Ala Ile Gly Glu Lys Glu Ala Arg Ala Ala Gln Ala Glu
 545 550 555 560

Glu Arg Lys Phe Gln Gln His Ile Leu Gly Gln Gln Lys Lys Glu Leu
 50 565 570 575

Ala Ala Leu Leu Glu Ala Gln Lys Arg Thr Tyr Lys Leu Arg Lys Glu
 580 585 590

55 Gln Leu Lys Glu Glu Leu Gln Glu Asn Pro Ser Thr Pro Lys Arg Glu
 595 600 605

Lys Ala Glu Trp Leu Leu Arg Gln Lys Glu Gln Leu Gln Gln Cys Gln
 610 615 620

Ala Glu Glu Glu Ala Gly Leu Leu Arg Arg Gln Arg Gln Tyr Phe Glu
 5 625 630 635 640

Leu Gln Cys Arg Gln Tyr Lys Arg Lys Met Leu Leu Ala Arg His Ser
 645 650 655

10 Leu Asp Gln Asp Leu Leu Arg Glu Asp Leu Asn Lys Lys Gln Thr Gln
 660 665 670

Lys Asp Leu Glu Cys Ala Leu Leu Arg Gln His Glu Ala Thr Arg
 15 675 680 685

Glu Leu Glu Leu Arg Gln Leu Gln Ala Val Gln Arg Thr Arg Ala Glu
 690 695 700

Leu Thr Arg Leu Gln His Gln Thr Glu Leu Gly Asn Gln Leu Glu Tyr
 20 705 710 715 720

Asn Lys Arg Arg Glu Gln Glu Leu Arg Gln Lys His Ala Ala Gln Val
 725 730 735

25 Arg Gln Gln Pro Lys Ser Leu Lys Ser Lys Glu Leu Gln Ile Lys Lys
 740 745 750

Gln Phe Gln Glu Thr Cys Lys Ile Gln Thr Arg Gln Tyr Lys Ala Leu
 30 755 760 765

Arg Ala His Leu Leu Glu Thr Thr Pro Lys Ala Gln His Lys Ser Leu
 770 775 780

Leu Lys Arg Leu Lys Glu Glu Gln Thr Arg Lys Leu Ala Ile Leu Ala
 35 785 790 795 800

Glu Gln Tyr Asp Gln Ser Ile Ser Glu Met Leu Ser Ser Gln Ala Leu
 805 810 815

40 Arg Leu Asp Glu Thr Gln Glu Ala Glu Phe Gln Ala Leu Arg Gln Gln
 820 825 830

Leu Gln Gln Glu Leu Glu Leu Asn Ala Tyr Gln Ser Lys Ile Lys
 835 840 845

45 Ile Arg Thr Glu Ser Gln His Glu Arg Glu Leu Arg Glu Leu Glu Gln
 850 855 860

Arg Val Ala Leu Arg Arg Ala Leu Leu Glu Gln Arg Val Glu Glu Glu
 50 865 870 875 880

Leu Leu Ala Leu Gln Thr Gly Arg Ser Glu Arg Ile Arg Ser Leu Leu
 885 890 895

55 Glu Arg Gln Ala Arg Glu Ile Glu Ala Phe Asp Ala Glu Ser Met Arg
 900 905 910

Leu Gly Phe Ser Ser Met Ala Leu Gly Gly Ile Pro Ala Glu Ala Ala
915 920 925

5 Ala Gln Gly Tyr Pro Ala Pro Pro Pro Ala Pro Ala Trp Pro Ser Arg
930 935 940

Pro Val Pro Arg Ser Gly Ala His Trp Ser His Gly Pro Pro Pro Pro
945 950 955 960

10 Gly Met Pro Pro Pro Ala Trp Arg Gln Pro Ser Leu Leu Ala Pro Pro
965 970 975

Gly Pro Pro Asn Trp Leu Gly Pro Pro Thr Gln Ser Gly Thr Pro Arg
980 985 990

15 Gly Gly Ala Leu Leu Leu Leu Arg Asn Ser Pro Gln Pro Leu Arg Arg
995 1000 1005

20 Ala Ala Ser Gly Gly Ser Gly Ser Glu Asn Val Gly Pro Pro Ala
1010 1015 1020

Ala Ala Val Pro Gly Pro Leu Ser Arg Ser Thr Ser Val Ala Ser
1025 1030 1035

25 His Ile Leu Asn Gly Ser Ser His Phe Tyr Ser
1040 1045

30 <210> 9

<211> 3285

<212> DNA

35 <213> Homo sapiens

40 <220>

<221> misc_feature

<222> (3245)..(3245)

45 <223> a, t, c or g

50 <220>

<221> misc_feature

<222> (3263)..(3263)

<223> a, t, c or g

55 <220>

5 <221> misc_feature
 <222> (3278)..(3278)
5 <223> a, t, c or g

10 <220>
10 <221> misc_feature
 <222> (72)..(72)
15 <223> a, t, c or g
15
 <300>
20 <308> EMBL AY049015
20 <309> 2001-10-15
 <313> (1)..(3285)

25 <400> 9
 caacggattt catttcatac agatgaacca aggatcggga tagcagtata aaattagaat 60
30 caagacagct gnctgccaaag caggatgcc a tcaactaaca ggcaggcag cctgaaggac 120
 cctgaaattt cagagcttt cttcaaagaa gatccagaga agctttcac agatctcaga 180
 gaaattggcc atggaagctt tggagcagtg tattttgcac gagatgtgcg taccaatgaa 240
35 gtggtggcca tcaagaaaat gtcttatagt ggaaagcagt ctactgagaa atggcaggat 300
 attattaagg aagtcaagtt tctacaaaga ataaaacatc ccaacagtat agaatacaaa 360
40 ggctgttatt tacgtgaaca cacagcatgg cttgtaatgg aatattgttt aggatctgct 420
 tcggattttac tagaagttca caaaaagcca ttacaagaag tggaaatagc agcaattaca 480
 catggtgctc ttcagggatt agcctactta cattctcata ctatgattca tagagatatc 540
45 aaaggaggaa atatccttct gacagaacca ggccagggtga aacttgctga ctttggctct 600
 gcttccatgg catcacctgc caattccctt gtggaaacgc cgtattggat ggccccagaa 660
50 gtaattttag ccatggatga aggacaatat gatggcaaag tagatgtgtg gtctttggaa 720
 ataacatgtt ttgaacttagc ggaaaggaag cctcctttat ttaatatgaa tgcaatgagt 780
 gccttatatc acatagccca aaatgaatcc cctacactac agtctaattga atggctgtat 840
55 tatttcgca actttgtaga ttcttgccctc cagaaaatcc ctcaagatcg acctacatca 900
 gaggaacttt taaagcacat atttgttctt cgggagcggcc ctgaaaccgt gttatagat 960

	ctcattcaga ggacaaagga tgcagtaaga gagctggaca atctgcagta tcgaaagatg	1020
5	aagaaaactcc ttttccagga ggcacataat ggaccagcag tagaagcaca ggaagaagaa	1080
	gaggaacaag atcatggtgt tggccggaca ggaacagtta atagtgttgg aagtaatcaa	1140
	tccattccca gcatgtccat cagtgccagc agccaaagca gtagtgttaa cagtcttcca	1200
10	gatgtctcag atgacaagag tgagctagac atgatggagg gagaccacac agtcatgtct	1260
	aacagttctg ttatccattt aaaaccagag gaagaaaatt acagagaaga gggagatcct	1320
15	agaacaagag catcagatcc acaatctcca ccccaagtat ctcgtcacaa atcacactat	1380
	cgtaatcgag aacactttgc tactatacgg acagcatcac tggttacgag gcaaattgcaa	1440
	gaacatgagc aggactctga gcttagagaa caaatgtctg gctataagcg aatgaggcga	1500
20	caacatcaaa agcaactgat gactctggaa aacaagctaa aggctgagat ggatgaacat	1560
	cgcctcagat tagacaaaga tcttggaaact cagcgtaaaca attttgc agaaatggag	1620
25	aaaccttatca agaaacacca ggctgctatg gagaagagg ctaaagtatgat gtccaatgaa	1680
	gagaaaaat ttcagcaaca tattcaggcc caacagaaga aagaactgaa tagtttctc	1740
	gagtcccaga aaagagagta taaacttcga aaagagcagc taaaagagga gctaaatgaa	1800
30	aaccagagta ccccaaaaaa agaaaaacag gagttggctt caaaggcagaa ggagaatata	1860
	cagcatttcc aagcagaaga agaagctaac cttcttcgac gtcaaagaca atacctagag	1920
35	ctggaatgcc gtcgcttcaa gagaagaatg ttacttgggc gtcataactt agagcaggac	1980
	cttgcaggagg aggagttaaa caaaagacag actcagaagg acttagagca tgccatgcta	2040
	ctccgacagc atgaatctat gcaagaactg gagttccgccc acctcaacac aattcagaag	2100
40	atgcgctgtg agttgatcat attacagcat caaactgagc tcactaacca gctggaatata	2160
	aataagcgaa gagaacgaga actaagacga aagcatgtca tggaagttcg acaacagcct	2220
45	aagagtttga agtctaaaga actccaaata aaaaagcagt ttcaaggatac ctgcaaaatc	2280
	caaaccagac agtacaaagc attaagaaat cacctgtgg agactacacc aaagagttag	2340
	cacaaagctg ttctgaaacg gctcaaggag gaacagacccc ggaaattagc tatcttggct	2400
50	gagcagtatg atcacagcat taatgaaatg ctctccacac aagccctgcg tttggatgaa	2460
	gcacaggaag cagagtgcctt ggtttgaag atgcagctgc agcaggaact ggagctgttg	2520
55	aatgcgtatc agagcaaaat caagatgcaa gctgaggcac aacatgatcg agagcttcgc	2580
	gagcttgaac agagggtctc cctccggagg gcactcttag aacaaaagat tgaagaagag	2640

atgttggctt tgcagaatga gcgcacagaa cgaatacgaa gcctgttgg aCGTCAAGCC 2700
 agagagattg aAGCTTTGA CTCTGAAAGC ATGAGACTAG GTTTAGTAA TATGGTGCTT 2760
 5 tctaattctct cccctgaggc attcagccac agctaccgg gagcttctgg ttggtcacac 2820
 aaccctactg ggggtccagg acctcaactgg ggtcatccca tgggtggccc accacaagct 2880
 10 tggggccatc caatgcaagg tggaccccag ccatgggtc acccttcagg gccaatgcaa 2940
 ggggtacctc gaggttagcag tatggagtc cgcaatagcc cccaagctct gaggcggaca 3000
 gcttctgggg gacggacgga gcagggcatg agcagaagca cgagtgtcac ttcacaaaata 3060
 15 tccaatgggt cacacatgtc ttatacataa cttataatt gagagtggca attccgctgg 3120
 agctgtctgc caaaagaaac tgcctacaga catcatcaca gcagcctcct cacttggta 3180
 20 ctacagtgtg gaagctgagt gcatatggta tattttattc atttttgtaa aCGTTCTGT 3240
 tttngttta ctaattggga tgncatagta cttggctncc cgggt 3285

25 <210> 10
 <211> 1005
 <212> PRT
 30 <213> Homo sapiens

<300>
 35 <308> SpTrEMBL
 <309> 2000-10-01
 <313> (1)...(1005)
 40

<400> 10

45 Leu Leu Ser Arg Met Pro Ser Thr Asn Arg Ala Gly Ser Leu Lys Asp
 1 5 10 15

Pro Glu Ile Ala Glu Leu Phe Phe Lys Glu Asp Pro Glu Lys Leu Phe
 20 25 30

50 Thr Asp Leu Arg Glu Ile Gly His Gly Ser Phe Gly Ala Val Tyr Phe
 35 40 45

Ala Arg Asp Val Arg Thr Asn Glu Val Val Ala Ile Lys Lys Met Ser
 50 55 60

55 Tyr Ser Gly Lys Gln Ser Thr Glu Lys Trp Gln Asp Ile Ile Lys Glu
 65 70 75 80

Val Lys Phe Leu Gln Arg Ile Lys His Pro Asn Ser Ile Glu Tyr Lys
 85 90 95

5 Gly Cys Tyr Leu Arg Glu His Thr Ala Trp Leu Val Met Glu Tyr Cys
 100 105 110

Leu Gly Ser Ala Ser Asp Leu Leu Glu Val His Lys Lys Pro Leu Gln
 115 120 125

10 Glu Val Glu Ile Ala Ala Ile Thr His Gly Ala Leu Gln Gly Leu Ala
 130 135 140

Tyr Leu His Ser His Thr Met Ile His Arg Asp Ile Lys Ala Gly Asn
 145 150 155 160

Ile Leu Leu Thr Glu Pro Gly Gln Val Lys Leu Ala Asp Phe Gly Ser
 165 170 175

20 Ala Ser Met Ala Ser Pro Ala Asn Ser Phe Val Gly Thr Pro Tyr Trp
 180 185 190

Met Ala Pro Glu Val Ile Leu Ala Met Asp Glu Gly Gln Tyr Asp Gly
 195 200 205

25 Lys Val Asp Val Trp Ser Leu Gly Ile Thr Cys Ile Glu Leu Ala Glu
 210 215 220

Arg Lys Pro Pro Leu Phe Asn Met Asn Ala Met Ser Ala Leu Tyr His
 225 230 235 240

Ile Ala Gln Asn Glu Ser Pro Thr Leu Gln Ser Asn Glu Trp Ser Asp
 245 250 255

35 Tyr Phe Arg Asn Phe Val Asp Ser Cys Leu Gln Lys Ile Pro Gln Asp
 260 265 270

Arg Pro Thr Ser Glu Glu Leu Leu Lys His Ile Phe Val Leu Arg Glu
 275 280 285

40 Arg Pro Glu Thr Val Leu Ile Asp Leu Ile Gln Arg Thr Lys Asp Ala
 290 295 300

Val Arg Glu Leu Asp Asn Leu Gln Tyr Arg Lys Met Lys Lys Leu Leu
 45 305 310 315 320

Phe Gln Glu Ala His Asn Gly Pro Ala Val Glu Ala Gln Glu Glu
 325 330 335

50 Glu Glu Gln Asp His Gly Val Gly Arg Thr Gly Thr Val Asn Ser Val
 340 345 350

Gly Ser Asn Gln Ser Ile Pro Ser Met Ser Ile Ser Ala Ser Ser Gln
 355 360 365

55 Ser Ser Ser Val Asn Ser Leu Pro Asp Val Ser Asp Asp Lys Ser Glu
 370 375 380

Leu Asp Met Met Glu Gly Asp His Thr Val Met Ser Asn Ser Ser Val
 385 390 395 400

5 Ile His Leu Lys Pro Glu Glu Glu Asn Tyr Arg Glu Glu Gly Asp Pro
 405 410 415

Arg Thr Arg Ala Ser Asp Pro Gln Ser Pro Pro Gln Val Ser Arg His
 420 425 430

10 Lys Ser His Tyr Arg Asn Arg Glu His Phe Ala Thr Ile Arg Thr Ala
 435 440 445

Ser Leu Val Thr Arg Gln Met Gln Glu His Glu Gln Asp Ser Glu Leu
 15 450 455 460

Arg Glu Gln Met Ser Gly Tyr Lys Arg Met Arg Arg Gln His Gln Lys
 465 470 475 480

20 Gln Leu Met Thr Leu Glu Asn Lys Leu Lys Ala Glu Met Asp Glu His
 485 490 495

Arg Leu Arg Leu Asp Lys Asp Leu Glu Thr Gln Arg Asn Asn Phe Ala
 500 505 510

25 Ala Glu Met Glu Lys Leu Ile Lys Lys His Gln Ala Ala Met Glu Lys
 515 520 525

Glu Ala Lys Val Met Ser Asn Glu Glu Lys Lys Phe Gln Gln His Ile
 30 530 535 540

Gln Ala Gln Gln Lys Lys Glu Leu Asn Ser Phe Leu Glu Ser Gln Lys
 545 550 555 560

35 Arg Glu Tyr Lys Leu Arg Lys Glu Gln Leu Lys Glu Glu Leu Asn Glu
 565 570 575

Asn Gln Ser Thr Pro Lys Lys Glu Lys Gln Glu Trp Leu Ser Lys Gln
 580 585 590

40 Lys Glu Asn Ile Gln His Phe Gln Ala Glu Glu Ala Asn Leu Leu
 595 600 605

Arg Arg Gln Arg Gln Tyr Leu Glu Leu Glu Cys Arg Arg Phe Lys Arg
 610 615 620

Arg Met Leu Leu Gly Arg His Asn Leu Glu Gln Asp Leu Val Arg Glu
 625 630 635 640

50 Glu Leu Asn Lys Arg Gln Thr Gln Lys Asp Leu Glu His Ala Met Leu
 645 650 655

Leu Arg Gln His Glu Ser Met Gln Glu Leu Glu Phe Arg His Leu Asn
 660 665 670

55 Thr Ile Gln Lys Met Arg Cys Glu Leu Ile Arg Leu Gln His Gln Thr
 675 680 685

Glu Leu Thr Asn Gln Leu Glu Tyr Asn Lys Arg Arg Glu Arg Glu Leu
 690 695 700

5 Arg Arg Lys His Val Met Glu Val Arg Gln Gln Pro Lys Ser Leu Lys
 705 710 715 720

Ser Lys Glu Leu Gln Ile Lys Lys Gln Phe Gln Asp Thr Cys Lys Ile
 725 730 735

10 Gln Thr Arg Gln Tyr Lys Ala Leu Arg Asn His Leu Leu Glu Thr Thr
 740 745 750

Pro Lys Ser Glu His Lys Ala Val Leu Lys Arg Leu Lys Glu Glu Gln
 15 755 760 765

Thr Arg Lys Leu Ala Ile Leu Ala Glu Gln Tyr Asp His Ser Ile Asn
 770 775 780

20 Glu Met Leu Ser Thr Gln Ala Leu Arg Leu Asp Glu Ala Gln Glu Ala
 785 790 795 800

Glu Cys Gln Val Leu Lys Met Gln Leu Gln Gln Glu Leu Glu Leu Leu
 805 810 815

25 Asn Ala Tyr Gln Ser Lys Ile Lys Met Gln Ala Glu Ala Gln His Asp
 820 825 830

Arg Glu Leu Arg Glu Leu Glu Gln Arg Val Ser Leu Arg Arg Ala Leu
 30 835 840 845

Leu Glu Gln Lys Ile Glu Glu Glu Met Leu Ala Leu Gln Asn Glu Arg
 850 855 860

35 Thr Glu Arg Ile Arg Ser Leu Leu Glu Arg Gln Ala Arg Glu Ile Glu
 865 870 875 880

Ala Phe Asp Ser Glu Ser Met Arg Leu Gly Phe Ser Asn Met Val Leu
 885 890 895

40 Ser Asn Leu Ser Pro Glu Ala Phe Ser His Ser Tyr Pro Gly Ala Ser
 900 905 910

Gly Trp Ser His Asn Pro Thr Gly Gly Pro Gly Pro His Trp Gly His
 45 915 920 925

Pro Met Gly Gly Pro Pro Gln Ala Trp Gly His Pro Met Gln Gly Gly
 930 935 940

50 Pro Gln Pro Trp Gly His Pro Ser Gly Pro Met Gln Gly Val Pro Arg
 945 950 955 960

Gly Ser Ser Met Gly Val Arg Asn Ser Pro Gln Ala Leu Arg Arg Thr
 965 970 975

55 Ala Ser Gly Gly Arg Thr Glu Gln Gly Met Ser Arg Ser Thr Ser Val
 980 985 990

Thr Ser Gln Ile Ser Asn Gly Ser His Met Ser Tyr Thr
995 1000 1005

```

5      <210>  11
6      <211>  1576
7      <212>  DNA
8      <213>  artificial sequence
9
10     <220>
11
12     <223>  RNAi fragment T17E9.1a (kin-18)
13
14     <400>  11
15     cgaaaaccag cagaagagcg aatgtcagct gaagaatgct ttcgacatcc attcattcaa 60
16     cggtctcgcc catcagacac aattcaggaa ctcattcaga gaacgaaaaa tatggattta
17     gagttggata atttcaata caaaaagatg agaaaactca tgtatggaa tgaaacagaa 120
18     gaaaaagaag gaagtgaagg aaatggagca tctgatgatt tagatttca tggaaatgaa 180
19     gctaattcaa ttggaagagg tagttttaa aattcaaagt gaaaatatta atatcttggaa 240
20     ataattttta taatattgct taaaaccctc agctttttt tgcagactct atcccttagt 300
21     tgttcgtttt ccatctattc tcgttttcag caggagattc tgcgtcatct cgaagtgcct 360
22     ctcttacttc ttccgatca atgcagagta gtggaggagc tggctttta gtgtccacca 420
23     atacgacggg tgctatggat aatgtgcatt gtactgtact gtttttttgg ttttaggaat 480
24     ggctttatta ttccctgcaa agttcaaaaa ttccatttat ttagttttt ctctcgaaat 540
25     tcatcgcgca acattgagaa tctttcaaaa tttcaggat cctctggata cggtaatggaa 600
26     agtagttcga cgacgagctc cgacgcccgc cgtcctccaa ttccctcgca aatgctct 660
27     tctacatcaa cgtctggtgt tggaaactatg ccgagtcatt gatcagttgg agcatcgatt 720
28     acggcgatcg cagtcaatcc aacaccgtct ctttcagaac ctatccaaac atcacaacca 780
29     acatcgaaat cagaatcatc ttctatactc gaaactgcac acgatgatcc tttggacacg 840
30     tcgatacgtg ctccagtgaa agacttgcatt atgcccgcatt gagcagtcaa ggaacgaata 900
31     gccacgttgc aaaatcacaa attcgacg cttcggtttcc agagaataat caatcaggaa 960
32     caagaagaat atacgaaaga gaacaatatg tatgagcaaa tgagcaagta caagcatcta 1020
33     cgacaaagcac atcacaacaa gctccaaacaa ttgaaagaac gatgtgcatt agatagagag 1080
34     1140

```

caactgcgtg tgaaaatgga tcgagaactc gaacaattga caacgacata ctcgaaagaa 1200
aagatgagag tgaggtgttc acagaataat gaactagaca aacggaaaaa agatatcgaa 1260
5 gatggggaga aaaagatgaa aaagacgaaa aatagtcaaa atcagcagca gatgaaactg 1320
tattcagcgc aacaattgaa agaatacaag tataacaagg aggcacagaa aacagtgaga 1380
10 attcacttt atttgatttc tgtaaagaaa ttatacgaaa ttttagacttt ataaaatttt 1440
aaatatgaaa gttctggtca cttttcagc tgcttctcca ctttttcaa agtttattat 1500
ttagtcttga ataattttt aaaaaatgtc ctaaaccaag aattttcagc gattacgaag 1560
15 tctgaacatg cctcga 1576

<210> 12
20 <211> 20
<212> DNA
25 <213> artificial sequence
<220>
30 <223> primer
<400> 12
cgaaaaccag cagaagagcg 20
35 <210> 13
<211> 21
<212> DNA
40 <213> artificial sequence
<220>
45 <223> primer
<400> 13
tcgaggcatg ttcagacttc g 21
50